An Experimental Framework to Examine the Influence of Promoter Architecture and Genomic Context on Gene Expression

DISSERTATION

zur Erlangung des Grades eines
Doktor der Naturwissenschaften
(Dr. rer. nat.)

Dem Fachbereich der Biologie
der Philipps-Universität Marburg
vorgelegt

von
CARLO ANTON KLEIN

Master of Science
Aus Kassel

Marburg, 2020

Die Untersuchungen zur vorliegenden Arbeit wurden von Februar 2016 bis Dezember 2019 unter der Betreuung von Dr. Patrick Sobetzko und Prof. Dr. Torsten Waldminghaus am LOEWE-Zentrum für Synthetische Mikrobiologie in der Arbeitsgruppe Sobetzko in Marburg durchgeführt.

Vom Fachbereich Biologie

der Philipps-Universität Marburg als Dissertation

angenommen am:

Erstgutachter: Dr. Patrick Sobetzko

Zweitgutachter: Prof. Dr. Torsten Waldminghaus

Tag der Disputation:

Summary

Transcription is a fundamental process of gene expression. Information stored in DNA is transcribed into different types of mobile RNA, which play a role in various essential processes of the cell, e.g. translation. However, cells do not need all the information stored in their DNA at the same time. Therefore, the process of transcription gets regulated by a plethora of mechanisms. One frequently discussed but poorly understood mechanism of transcription regulation is DNA supercoiling [Travers and Muskhelishvili, 2005]. Whereby, the process of transcription itself affects the DNA-topology up- and downstream of the transcription machinery as described in the twin supercoiling domain model [Liu and Wang, 1987]. This phenomenon is called Transcription Coupled DNA Supercoiling (TCDS). It has also been shown that genes react individually to changes in DNA supercoiling and that there is a selection pressure on adapting to the DNA supercoiling levels emitted by neighbouring gene expression [Sobetzko, 2016]. The system in which promoters react to changes in DNA supercoiling is as diverse as there are promoters; notably, some promoters seem not to respond to DNA supercoiling at all. Thus, this raises the question as to which elements within different promoter types cause them to respond to TCDS so differently.

In this thesis, I built a pipeline to investigate the effects of TCDS and DNA supercoiling on promoters. Firstly, I created a plasmid toolbox, which allows modular assembly of transcription units. The central feature of this toolbox is the flexibility to test different arrangements of multiple transcription units. I achieved this by adapting the well established Modular Cloning (MoClo) standard [Weber et al., 2011] and build my toolbox around it. I thus created a system that works on both its own and is compatible with the existing standard MoClo protocol.

In the second part of this thesis, I established an experimental pipeline using synthetic σ^{70}-promoters to investigate the influence of DNA supercoiling on transcription. The experimental setup allowed precise changes in parts of the promoter and at the same time created a library of these promoters. Using this pipeline to investigate the spacer region of the promoter, I was able to confirm that the spacer influences the promoter strength. Further, I showed that the promoter spacer has only a limited effect on the supercoiling sensitivity of a promoter. I also showed that a 5'-TGTG-3' motif in the spacer region could lower transcription by enhancing RNA-polymerase (RNAP)-binding. Moreover, the experimental setup also showed the constraints of using the DNA-relaxing drug novobiocin on a plasmid-based system. Hence, to further investigate the effects of TCDS on neighbouring transcription, I applied an optogenetically-controllable promoter to the previously established pipeline.

Finally, I began to explore the possibility of integrating my experimental promoter setup into any genomic position. As such, a CRISPR/Cas9-based homologous recombination system was developed further to make it modular and compatible with the Modular Cloning protocol. I could show the first features of this system to work.

Zusammenfassung

Die Transkription ist ein grundlegender Prozess der Genexpression. Die in der DNA gespeicherten Informationen werden in mobile RNA transkribiert, die bei verschiedenen essenziellen Prozessen der Zelle wie z.B. der Translation, eine Rolle spielen. Die Zellen benötigen jedoch nicht alle in der DNA gespeicherten Informationen gleichzeitig. Daher wird der Prozess der Transkription durch eine Vielzahl von Mechanismen reguliert. Ein häufig diskutierter, aber nicht vollständig verstandener Transkriptionsregulator ist das DNA-*Supercoiling* [Travers and Muskhelishvili, 2005]. Wobei der Transkriptionsprozess selbst die DNA-Topologie vor und hinter der Transkriptionsmaschinerie beeinflusst, wie im *twin supercoiling domain*-Modell von Liu and Wang [1987] beschrieben. Dieses Phänomen wird als Transkriptions-gekoppeltes DNA-*Supercoiling* (TCDS) bezeichnet. Darüber hinaus reagieren die Gene individuell auf Veränderungen des DNA-*Supercoilings* und es besteht ein Selektionsdruck für Gene sich an die DNA-*Supercoiling*-Niveaus, welche durch das benachbarte Expressionsverhalten emittiert werden, anzupassen [Sobetzko, 2016]. Die Reaktionen der Promotoren auf Veränderungen im DNA-*Supercoiling* sind so vielfältig wie die Promotoren selbst. Insbesondere scheinen einige Promotoren nicht auf Änderungen im DNA-*Supercoiling* zu reagieren. Dies führt zu der Frage was die Vielfalt der Promotoren dazu bringt so unterschiedlich auf TCDS zu reagieren.

Im Laufe dieser Arbeit wurde ein Versuchsprozess entwickelt, um die Auswirkungen von DNA-*Supercoiling* und TCDS auf Promotoren zu untersuchen. Zuerst wurde dafür eine Plasmid-Toolbox erstellt, die den modularen Aufbau von Transkriptionseinheiten ermöglicht. Das zentrale Merkmal dieser Toolbox ist die Flexibilität, verschiedene Anordnungen derselben Teile zu testen. Dies wurde erreicht, indem der gut etablierte MoClo-Standard adaptiert wurde und die Toolbox um diesen herum aufgebaut wurde. Dadurch wurde ein System geschaffen, welches eigenständig

funktioniert, mit dem bestehenden MoClo-Standard kompatibel ist und ihn erweitert.

Im zweiten Teil dieser Arbeit wurde ein Experimentaufbau entwickelt in welchem mit Hilfe von synthetischen σ^{70}-Promotoren, der Einfluss von DNA-*Supercoiling* auf die Transkription untersucht werden kann. Der Versuchsaufbau ermöglichte präzise Veränderungen in Teilen des Promotors und schaffte gleichzeitig eine *library* dieser Promotoren. Mit Hilfe dieser Experimente zur Untersuchung der *spacer*-Region des Promotors konnte bestätigt werden, dass der *spacer* die Stärke des Promotors beeinflusst. Außerdem konnte gezeigt werden, dass der *spacer* die *Supercoiling*-Empfindlichkeit nur geringfügig beeinflusst. Ferner konnte gezeiget, dass ein 5‘-TGTG-3‘-Motiv im *spacer* die Transkription durch eine verstärkte RNAP-Bindung senken könnte. Der Versuchsaufbau zeigte jedoch die Einschränkungen bei der Verwendung der DNA-relaxierenden Chemikalie Novobiocin mit einem plasmidbasierten System. Deshalb – und um die Auswirkungen von TCDS auf die benachbarte Transkription weiter zu untersuchen, wurde ein optogenetisch kontrollierbarer Promotor in den bereits etablierten Versuchsaufbau integriert.

Schließlich wurden die ersten Schritte unternommen, um einen Weg zu finden, den Promoter-Test in jede beliebige genomische Stelle integrieren zu können. Dafür wurde ein auf CRISPR/Cas9 basierendes, homologes Rekombinationssystem weiterentwickelt und modularisiert, welches DNA-Fragmente nahtlos ins Genom integrieren kann.

Contents

List of Acronyms

AGE Agarose Gel Electrophoresis 67
APS ammonium persulphate solution 72
BSA Bovine serum albumin 72
CDS coding sequence 4
crRNA CRISPR-RNA 1
CTD C-terminal domain 2
DMSO dimethyl sulfoxide 68
DNA deoxyribonucleic acid 68
dNTP deoxyribose nucleoside triphospates 66
dsDNA double-stranded DNA 28
DTT Dithiothreitol 79
e.g. *exempli gratia* 66
E. coli *Escherichia coli* 4
EDTA Ethylenediaminetetraacetic acid 78
EMSA Electrophoretic Mobility Shift Assay 71
etc *et cetera* 76
IPTG isopropyl β-d-1-thiogalactopyranoside 28
LB Lysogeny Broth 73
MoClo Modular Cloning i
MoCloFlex Flexible Modular Cloning 17
mRNA messenger RNA 1
nt nucleotide 27
ORF open reading frame 27
ori origin of replication 27
PAGE Polyacrylamide Gel Electrophoresis 72
PCR Polymerase Chain Reaction 17

RBS ribosome binding site 16
RNAP RNA-polymerase ii
rRNA ribosomal RNA 1
TAE Tris-acetate-EDTA 79
Taq *Thermus aquaticus* 68
TBE TRIS-Borat-EDTA-Buffer 67
TBE Tris-borate-EDTA 67
TCDS Transcription Coupled DNA Supercoiling i
TEC termination elongation complex 2
TEMED Tetramethylethylendiamin 72
TFB1 Transformation Buffer 1 73
TFB2 Transformation Buffer 2 73
tracrRNA trans-activating crRNA 1
tRNA transfer RNA 1
X-gal 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside 28

1 Introduction

1.1 Transcription is the Start of Gene Expression

Transcription is a fundamental process in gene expression. According to the central dogma of molecular biology, it is one of the general sequential information flows in biology [Crick, 1970]. Transcription describes the process of how a ribonucleic acid (RNA) molecule is synthesised complementary to a deoxyribonucleic acid (DNA) template by an enzyme class called DNA dependent RNA-polymerases (RNAP). In other words, certain information stored on the stable and large molecules of DNA gets transcribed into the shorter, unstable and more mobile units of RNA. While eukaryotes may encode multiple RNAP subtypes, there is only one bacterial RNAP. The core (E) of this enzyme consists of two α, one β, one β', and one ω subunit. Together with a sigma factor (σ) the core forms the RNAP-holoenzyme (Eσ), which is necessary to start transcription. This holoenzyme is 449 kDa in size and has its active, central RNA-synthetase site built by the two subunits β and β' [Finn et al., 2000].

The bacterial RNAP produces all cellular RNAs. Together, these RNAs produced by transcription drive many essential processes in the cell. To name just a few, messenger RNA (mRNA) is the template for the translation of proteins [Brenner et al., 1961] , ribosomal RNA (rRNA) is essential for the formation of ribosomes [Urlaub et al., 1995], transfer RNA (tRNA) are important for the amino acid synthesis [Holley, 1965], and CRISPR-RNA (crRNA) as well as trans-activating crRNA (tracrRNA) are needed for

the bacterial immune system [Jinek et al., 2012]. This list of RNAs produced by transcription, far from being complete, already indicates the importance of transcription and its subsequent events for cellular life.

1.1.1 The Process of Prokaryotic Transcription can be Subdivided into Four Phases

Promoter binding The first step of transcription is binding of the RNAP-holoenzyme to a promoter. Thereby, the sigma factor subunit defines the specificity of the promoter interaction by interacting with two sequence motifs -35 and -10 bp upstream of the first transcribed base +1 (fig. 1.1 A). Further, the C-terminal domain (CTD) of the alpha subunits can interact with specific UP- elements placed further upstream. The UP-element is part of several regulation mechanisms [Ross et al., 1993; Browning and Busby, 2004]. When the RNAP-holoenzyme is bound to the promoter while the DNA is still double-stranded the resulting complex is referred to as 'closed' complex.

Initiation After binding to the promoter, the RNAP-holoenzyme catalyses the unwinding and opening of 10–14 bp DNA around the discriminator and parts of the -10 region of the promoter [Roberts and Roberts, 1996; Henderson et al., 2017]. This step is also called the open complex formation (fig. 1.1 B) [Saecker et al., 2011]. During initiation, the first short RNA molecules are synthesised; however, this is an abortive process, and few, short molecules are produced until the stable elongation starts.

Elongation Following the initial RNA synthesis, the sigma factor leaves the RNAP-holoenzyme and is free to bind another RNAP (fig. 1.1 C). The complex of RNAP (without sigma factor), nascent RNA, and the DNA is called the termination elongation complex (TEC). In its centre 10–14 bp of single-stranded DNA forms the so-called transcription bubble, that is travelling along the TEC. The RNA synthesis is catalysed in the active site formed by the two β subunits and has a mean speed of around 40–50 bp/s [Yamaguchi, 2013]. During elongation of the transcript, 2–3 turns of the DNA have to be opened, which creates asymmetric torsional stress on the DNA molecule. This torsional stress can lead to DNA supercoiling, which has to be released during

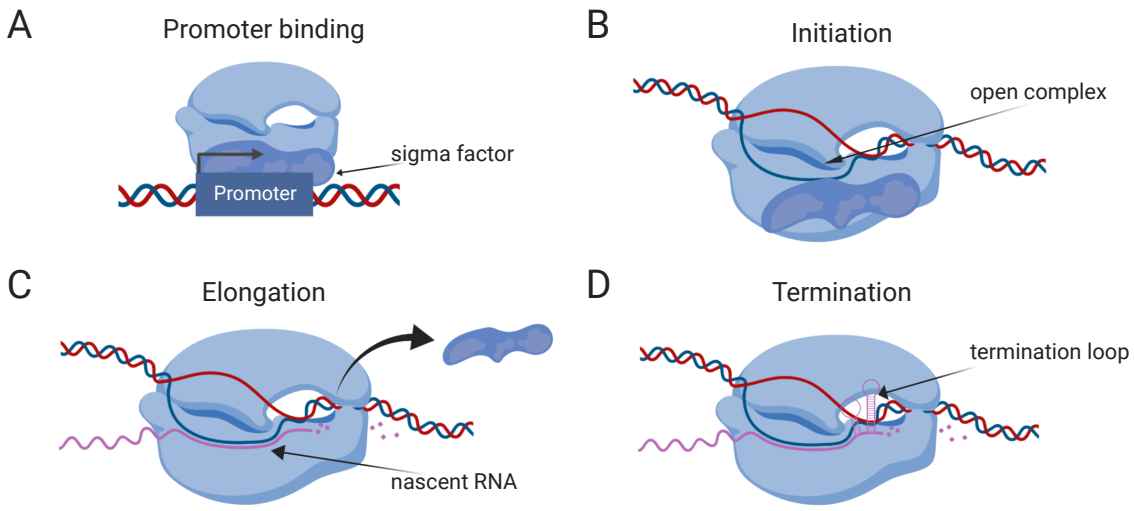


Figure 1.1: The four phases of the transcription process. **A** The assembled RNAP-holoenzyme (RNAP core plus sigma factor) bind the promoter. This promoter binding results in the closed complex. The unwinding and melting of the DNA and an often abortive RNA synthesis initiates the transcription **B**. After a stable RNA synthesis, the sigma factor dissolves, and the transcription elongation complex moves along the DNA **C**. One way to end transcription is the intrinsic termination in which the nascent RNA forms a stem-loop structure that, as a consequence, leads to the dissolving of the RNAP **D**.

the whole process [Liu and Wang, 1987]. Notably, due to the lack of a nucleus, translation takes place on the nascent RNA, coupling the TEC to numerous ribosomes and nascent amino acid chains. This coupling also prevents the R-loop formation between nascent RNA and the transcription bubble [Gowrishankar and Harinarayanan, 2004].

Termination Transcription ends, when the RNAP gets removed from the DNA either by facing a stem-loop formed at the nascent RNA stretch called a termination loop (intrinsic termination; fig. 1.1 D). This termination loop, together with a TEC associated protein NusA, stalls the RNAP on the DNA and a following uracil-rich stretch destabilises the TEC consequently causing the dissociation of the RNAP [Farnham and Platt, 1981; Wilson and von Hippel, 1995] Alternatively, rho-dependent termination occurs via Rho proteins forcibly removing the RNAP from the DNA (Rho-dependent termination) [Bogden et al., 1999].

1.2 Transcription is a Regulated Process

Responding to environmental changes is essential for bacteria to survive, grow, and proliferate. Different environments cause different growing conditions. Cells facing stress, like starvation, need other genes to be expressed than cells growing in a bacterial land of plenty. In order to adapt to stressful conditions long-term, one possibility for cells is to adjust their gene expression via the regulation of transcription. These long-term adjustments in transcription alter the abundance of mRNA levels and thus the possibility of subsequent events in gene expression. Prokaryotes can modulate transcription in multiple ways, either by changing the affinity of the holoenzyme to the promoter or by preventing the binding to the promoter. One class of master regulators of transcription are the sigma factors that, together with specific promoters, are necessary for transcription.

1.2.1 Promoters Decide Where Transcription Starts

The decision of which part of a DNA gets transcribed by RNAP is made at the promoter, which is a specific DNA sequence in close vicinity to the first transcribed base of a gene. Basepairs of the promoter are numbered -1 in bp distance to the first transcribed base (+1); as such, there is no 0. Generally, a promoter consists of sequence stretches that interact with the subunits of the RNAP-holoenzyme. The function of promoters is to enhance the probability of RNAP-holoenzymes binding a particular region of the genome. Since promoters are specific for certain sigma factors, their interaction thus also influence the timing of transcription. Promoters in *Escherichia coli* (*E. coli*) are usually located within the first 300 bp upstream of the first transcribed base of a coding sequence (CDS). However, some genes or operons can have more than one promoter, that can be specific for different sigma factors.

Nevertheless, the promoter structure has some defined areas relative to the first transcribed base (+1): the discriminator, -10-region, the spacer, -35-region and an

UP-element. The -10 region, together with the -35-sequence and parts of the spacer, interacts with the sigma factor of the RNA-polymerase holoenzyme. The consensus sequence for the -10 region is 3'-TATAAT-5' [Hawley and McClure, 1983]. However, this represents the consensus of the -10 that is bound by the housekeeping sigma factor σ^{70}. The -35 region of σ^{70} promoters has the consensus sequence 3'-TTGACA-5' [Hawley and McClure, 1983]. Between the -10 and the -35 region, is the so-called spacing region, also called the spacer. In *E. coli* promoters, the spacer is between 15 and 21 bp long. The most abundant (optimal) spacer is 17 bp $\pm$ 1 bp in length. For a long time, it was thought that the spacer was unimportant in promoter binding since the bases of the spacer are rarely conserved. However, in the 1980s, a conserved TG-motif at position -16 was found in gram-negative as well as in gram-positive bacteria. In *E. coli*, this motif occurs in roughly 25% of the promoters and, due to its proximity to the -10 region, was named the 'extended -10' [Burr et al., 2000]. Mainly promoters with no or a weak -35-sequence had this TG-motif, and thus it was inferred that the TG-motif compensates for the lack of the -35 region [Kumar et al., 1993]. Further upstream of the -35 region, an AT-rich area of around 40 bp was shown [Estrem et al., 1998]. This so-called 'UP-element' can interact with the alpha subunit of the RNAP and thus can increase the affinity of RNAP to the promoter [Ross et al., 1993]. The discriminator is the space between the first transcribed base and the -10-region and is responsible for influencing the melting of the DNA-double-helix during open-complex formation. The discriminator is usually 6 bp $\pm$ 1 bp long. Together, the described promoter elements attract the RNAP-holoenzyme and thereby modulate the ensuing expression strength of the downstream coding sequence. That promoters can have various combinations of promoter elements of different 'strength' which interact to form an individual promoter's strength is described in the so-called mix and match model [Hook-Barnard and Hinton, 2007].

1.2.2 Sigma-Factors Recognise Specific Promoters

Sigma factors are a class of prokaryotic proteins that can be compared to general transcription factors in eukaryotes like TFIIB; however, sigma factors do not bind DNA on their own but in complex with the RNAP. Together with the RNAP core enzyme

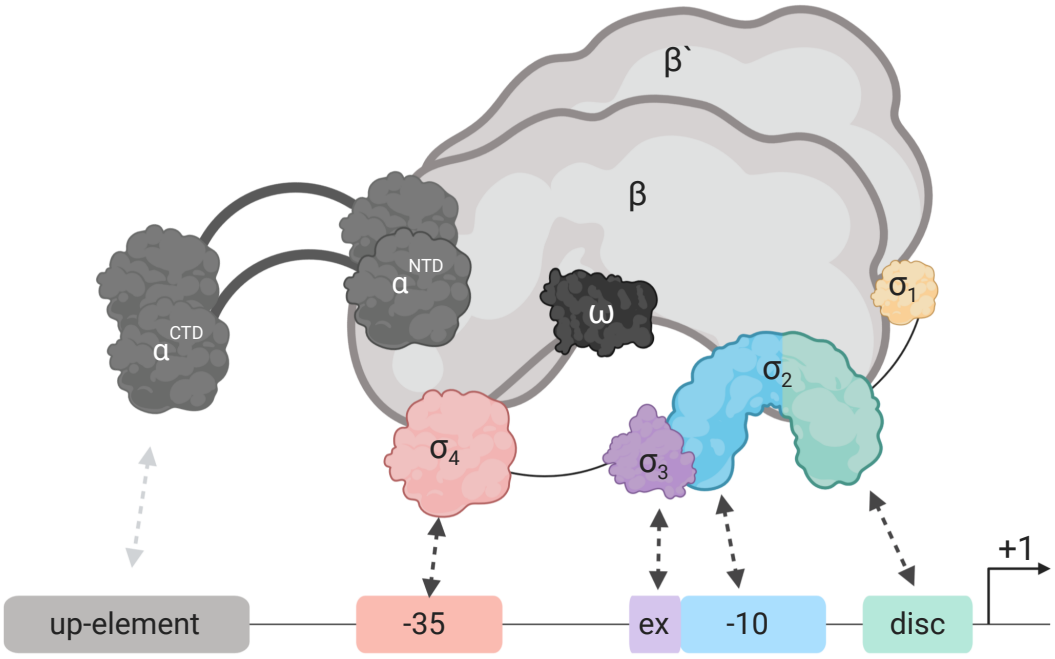


Figure 1.2: All subunits of the RNAP core enzyme (E, shades of grey) together with the σ^{70}-subunit (σ, coloured domains) form the RNAP-holoenzyme (Eσ). The σ_1-domain does not interact with the promoter, whereas the σ_2-domain interacts both with the discriminator and the -10 region of the promoter. σ_3 interacts with the so-called extended -10 motif and σ_4 with the -35 region (Adapted from [Davis et al., 2017]).

(E), sigma factors form the RNAP-holoenzyme (Eσ; fig. 1.2). Different classes of sigma factors, thereby, reversibly bind to the RNAP, giving the enzyme specificity to different promoters. Sigma factors can thus be seen as master regulators in transcription as all sigma factor classes together initiate the transcription of every gene in prokaryotes [Davis et al., 2017]. In *E. coli*, seven different sigma factors recognise cognate promoter sequences and ensure that these promoters are active when the expression of certain genes is essential for the cell. When cells grow exponentially, the most abundant sigma factor is the 70kDa RpoD, accordingly also called σ^{70}. σ^{70} controls the majority of *E. coli* genes and almost all genes necessary for growth and is therefore known as the housekeeping sigma factor [Dombroski et al., 1992]. When cells are stressed, and about to enter stationary phase, the sigma factor RpoS (σ^{38}) gets upregulated. RpoS

is essential for the general stress response of *E. coli* encompassing scenarios such as when toxic metabolites accumulate, the carbon sources are limited, or the cell density is high [Hengge-Aronis, 2002]. The other five sigma factors are for specific stress conditions or cell motility. RpoN (σ^{54}) is active when cells face nitrogen starvation, RpoH (σ^{32}) and RpoE (σ^{24}) are active in heat shock response, RpoF (σ^{28}) is essential for expressing cell motility genes, and FecI (σ^{19}) controls the iron transport. The central role of sigma factors in transcription and during stress response makes themselves a target for regulation. Regulation of sigma factor activity is mediated by anti-sigma factors which in turn bind a specific sigma factor, e.g. in the case of σ^{70} Rsd can prevent σ^{70} from binding RNAP, Rsd is itself antagonised by Hpr [Park et al., 2013]. This form of inhibition is also used as a tool in synthetic biology, e.g. for constructing orthogonal regulated gene circuits [Bervoets and Charlier, 2019]. Other forms of regulation are *inter alia* adjustments of sigma factor translation [Balandina et al., 2001] or can involve alarmons like ppGpp [Spira et al., 2008].

The σ^{70} has four domains that interact with distinct DNA sequences of the promoter (fig. 1.2). This interaction is due to the affinity of the sigma subunits for distinct parts of the promoter. Subunit 4 (σ_4) of the sigma factor attaches to the -35 region, subunit 3 (σ_3) to the extended -10, and subunit 2 (σ_2) binds to the -10 and the discriminator region. Subunit 1 (σ_1) does not interact with the promoter but is likely involved in the formation of the RNAP-holoenzyme. As such, these interactions also influence transcription initiation, since binding to -10 and discriminator helps to melt this region and so helps to form the open complex. Depending on the promoter, the spacing between -10 and -35 can be crucial for RNAP binding since the distance between the sigma factor domains σ_2 and σ_4 favour a certain distance of the -10 and -35 region [Shultzaberger et al., 2007].

1.3 Transcription modifies the local DNA supercoiling and *vice versa*

Transcription is one of the primary sources of DNA supercoiling in the cell, and more than half of the genes in *E. coli* are sensitive to changes in DNA supercoiling Blot et al.

[2006]. According to the Liu and Wang twin-supercoiled-domain model, transcription activity alters DNA supercoiling around the transcription complex in an asymmetric manner [Liu and Wang, 1987]. During transcription, the template DNA faces torsional stress. As the RNAP is a massive protein complex, and the nascent RNA is linked to it, the RNAP cannot move around the DNA helix; as a consequence, the DNA has to unwind into the complex. This process leads to overwinding downstream and underwinding upstream of the RNAP/DNA complex by changing the twists around the RNAP complex. Consequently, the DNA accumulates supercoiling up- and downstream of the transcription site (fig. 1.3). Work by Kouzine et al. has shown that this transcription-dependent supercoiling transmits 10–15 kb up- and downstream of the transcription site and is present across the genome [Kouzine et al., 2013]. TCDS is observed in prokaryotes as well as in eukaryotes [Meyer and Beslon, 2014] and seems to have similar effects on the chromatin organisation as reviewed in [Ma and Wang, 2016]. Thereby it seems to make no difference if the DNA template is circular or linear, since the rotation behaviour of a long linear DNA-molecule is similar to a molecule whose ends are fixated or fused [Kouzine et al., 2004, 2008]. Thus, neighbouring operons or genes face the supercoiling of ongoing transcription in their vicinity even when the full set of topoisomerases is present and active as is shown by [Lilley and Higgins, 1991]. As such, the supercoiling sensitivity of those promoters is dependent on the supercoiling resulting from transcription of nearby genomic regions.
Further, operons are ordered within the chromosome according to their preferred supercoiling sensitivity and the Ori-Ter supercoiling gradient [Sobetzko, 2016]. Negative supercoiling makes it easier to melt DNA at AT-rich sequences, such as the promoter, and thus to open the double-stranded DNA for the initiation of transcription [Rifka et al., 2015]. The nuclear binding protein FIS prefers to bind a certain level of negatively supercoiled DNA and by binding and repressing the gyrase promoters can be seen as part of the DNA supercoiling homeostasis mechanisms of the cell [Schneider et al., 1999; Muskhelishvili and Travers, 2003]. Thus, promoter binding and transcription initiation are probably the main regulatory targets of DNA supercoiling.
Nevertheless, the question of how the promoter itself can be responsive to supercoiling remains. One promoter element that influences the topology of the constituent parts of the promoter which interacts with the RNAP is the spacer. A spacer length of 17 bp is placing the centres of the two motifs -35 and -10 almost two helical turns of B-DNA apart [Wang, 1979].

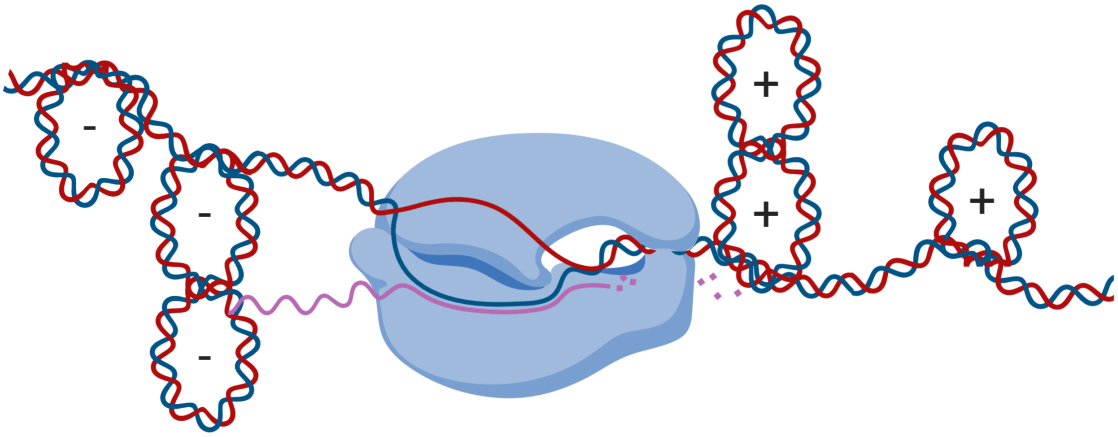


Figure 1.3: The twin supercoiled domain model. As the RNAP moves along the DNA template during transcription, the DNA is partially opened. Transcription, thereby, asymmetrically generates torsional stress. The DNA downstream is overwound and generates positive supercoils, and the DNA upstream is left underwound, generating negative supercoils [Liu and Wang, 1987].

Consequently, it was shown that the variation of the spacer length modulates the response of a promoter to DNA supercoiling [Aoyama and Takanami, 1988], thus making the spacer a *bona fide* target for TCDS research. However, experimental proof using more than one promoter at a time has until now been unavailable.

1.3.1 DNA Supercoiling and the Role of Topoisomerases

The universal carrier of genetic information, DNA, is a double-stranded nucleic acid, that consists of sequential combinations of the four nucleotides (G, A, T, C). The two strands are paired anti-parallel in their sequence. That means a G in the Crick strand pairs with a C in the Watson strand; the same is true for A pairing with T. This base pairing, together with the phosphate backbone of the strands, is forming the double-helical structure [Watson and Crick, 1953]. In this helical structure, both strands twist around each other with approximately 10.5 bp per turn. Thus, a DNA molecule with

10.5 bp/turn is called relaxed since it has no torsional stress. In *E. coli* and almost all bacteria, however, the circular DNA is not relaxed but negatively supercoiled. This implies that the number of helical turns called the linking number *Lk* is reduced [Deweese et al., 2008]. As an example, a DNA-molecule of 1050 bp length would have in its relaxed state a linking number Lk_0 (Eq. 1.3.1.1) of :

$$Lk_0 = 1050bp/10.5bp \tag{1.3.1.1}$$

$$Lk_0 = 100$$

The linking number represents the turns (also called twists, *Tw*) that are expected for a B-DNA molecule of a certain length. Notably, helical twists *Tw* can also transform into superhelical twists called writhes (*Wr*) and *vice versa*. The sum of twists and writhes is the linking number Lk of an existing DNA molecule (Eq. 1.3.1.2)

$$Lk = Tw + Wr \tag{1.3.1.2}$$

To determine if a DNA-molecule is overwound or underwound, meaning not in its relaxed state, ΔLk is calculated using equation Eq. 1.3.1.3.

$$\Delta Lk = Lk - Lk_0 \tag{1.3.1.3}$$

If ΔLk is > 0 , e.g. when $Lk = Tw + Wr = 101$, the DNA molecule has more helical twists or superhelical twists than expected in a relaxed molecule. Thus, the molecule is overwound or shows positive supercoiling. If ΔLk is < 0 , e.g. when $Lk = Tw + Wr = 99$, the DNA-molecule is underwound or negatively supercoiled. A standardised expression for supercoiling density independent of the length of the observed DNA molecule is the superhelical density σ, which can be calculated using equation Eq. 1.3.1.4.

$$\sigma = \Delta Lk/Lk_0 \tag{1.3.1.4}$$

As stated above, in *E. coli*, the chromosomes and plasmids are negatively supercoiled, which has several consequences. Firstly, supercoiled DNA physically occupies less space in the cell and thus, aids the compaction of the bacterial chromosome to fit into the cells [Higgins, 2016]. Secondly, underwound DNA requires less energy to melt, which then influences many cellular processes like transcription, replication, and re-

combination [Rifka et al., 2015; Muskhelishvili et al., 2016]. Nevertheless, nuances of superhelical density are far more complex. DNA supercoiling differs locally, responds to external stresses, and is negative during the growth cycle but slightly relaxes towards the stationary phase [Lal et al., 2016]. Further, supercoiling gets constrained by proteins such as NAPs or LacI [Muskhelishvili and Travers, 2003]. Recently it was shown that also replication initiation can be stopped by cells modulating their DNA supercoiling during stringent response [Kraemer et al., 2019].

Due to the central role of supercoiling in the cell, dedicated groups of enzymes called topoisomerases are responsible for controlling levels of supercoiling. Depending on the class of topoisomerases, they can introduce or relax DNA supercoiling. *E. coli* has four enzymes belonging to two types of topoisomerases: type IA and type IIA. Topoisomerases I and III belong to type IA. DNA gyrase and Topoisomerase IV belong to type IIA. The type IA topoisomerases act on negatively supercoiled DNA and can relax it; therefore, they do not need ATP. The type IIA enzymes Gyrase and Topo IV can, by consuming ATP, introduce negative supercoiling through inducing double-strand breaks followed by strand passage which alters the linking number in steps of two. However, Gyrase is mainly acting during transcription whereas Topo IV is needed to separate chromosomes during replication. (figure 1.4) [Champoux, 2001].

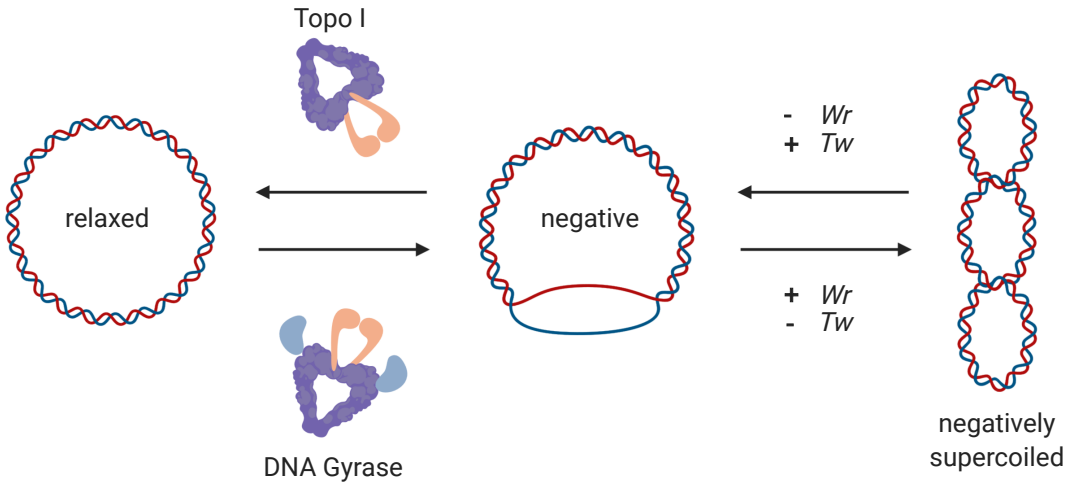


Figure 1.4: Schematic overview of the homeostatic action of Topo I and Gyrase. When the DNA has less than 10.5 bp/turn, it has fewer twists (*Tw*) than a relaxed molecule. Either the DNA is partially melted or faces torsional stress. By introducing one writhe (*Wr*), it can compensate one twist by generating a supercoiled structure and escape the torsional stress. Gyrase can transfer a positive supercoil into a negative supercoil by strand passage, changing the linking number in steps of -2. In the opposite direction Topo I can nick the DNA strand to reduce one twist, consequently relaxing the DNA.

1.4 Synthetic Biology: a Toolbox for Biological Research

Synthetic biology is a relatively young discipline in biological research. The word synthetic comes from the ancient Greek word συνθετικός (sunthetikós), meaning skilled in putting things together, or the art of composition. Further, the term synthetic is a synonym for artificial. The second term biology is composed of the ancient Greek word βιος (bios), which means life, and the ending -logy from logos an ending used to denote a specific field of scientific study. Thus, literally, synthetic biology describes the science of composing artificial life. However, synthetic biology is often difficult to discriminate from biotechnology, or molecular biology since these disciplines also manipulate biological parts to build biological systems to either produce something or to learn about the biological functions. Thus, both biotechnology and molecular biology create something artificial to answer their biological questions, which in turn

could allow these disciplines to be categorised as part of synthetic biology. This uncertainty is probably the reason why there are currently different definitions of synthetic biology. The definitions range from more biotechnological ones, like the one of the European Commission, to comprehensive ones dividing biology into either systematic or synthetic biology only [Schrauwers and Poolman, 2013]. The European Commission defined synthetic biology in 2005 as the "application of the engineering paradigm of systems designs to biology in order to produce predictable and robust systems with novel functionalities that do not exist in nature." This definition, like many others, refers to the standardisation and modularisation of biologic parts that are often seen as specific to synthetic biology [Keasling, 2006; Oldham et al., 2012]. From this viewpoint, synthetic biology is about applying engineering principles into molecular biology and focus on the engineering of biological systems. Synthetic biology as an engineering field in biology would thus solve specific problems rather than pursuing an understanding of the biological matter. However, creating or building something can also add to the understanding of the underlying biology. Thus, the key idea to test out biological 'bricks' in synthetic biology by modularisation of biological units such as promoters, terminators, combined transcription units, or small circuits and genetic switches will – besides producing libraries of these parts – also produce knowledge about the parts and their construction.

1.4.1 Modularisation Enhances the Cloning Process

Cloning is a fundamental method employed in the field of molecular biology. Regardless of the topic being investigated, most molecular biological experiments start with the cloning of plasmids or larger constructs. Cloning describes the process of creating and modifying specific genetic sequences including their amplification and transfer from one cell to another. It includes for example the adding of a tag to a protein, engineering a genetic circuit or even building large constructs like synthetic chromosomes. Thereby, some DNA parts are frequently used for different cloning reactions, such as antibiotic cassettes, origins of replication, tags or reporters. Hence, the repetitiveness and simplicity of a modular cloning approach can save both time and resources. Identifying the needs for systematic cloning, the synthetic biology

community has become one of the main drivers in the development of novel cloning systems [Ellis et al., 2011]. In parts, this is achieved by hierarchical cloning systems, which promote standardisation in the whole field of cloning. Methods like Golden Gate cloning, and its offshoot Modular Cloning (MoClo), reduce the needs for different restriction enzymes due to their reliance on type IIs restriction enzymes [Engler et al., 2008; Weber et al., 2011; Casini et al., 2015; Sarrion-Perdigones et al., 2013]. Since type IIs restriction enzymes cut outside their recognition site, they allow a rational design of overhangs. Additionally, using type IIs restriction enzymes allows the rearrangement of their restriction sites in such a way that they are removed when a cut vector and the insert ligate. Thus, restriction and ligation can be performed in a single step, making separate purification and ligation steps obsolete. During this so-called one-pot reaction, the final product gets enriched over time [Engler et al., 2008; Weber et al., 2011]. Following the idea of standardised and modularised cloning, a variety of methods appeared in the last years [Casini et al., 2015]. The list is also frequently being added to, e.g. by EcoFlex, MODAL, and PODAC [Van Hove et al., 2017; Moore et al., 2016; Storch et al., 2015]. Recent work using the MoClo protocol is dealing with the construction of synthetic chromosomes [Schindler et al., 2016; Messerschmidt et al., 2016; Zumkeller et al., 2018] or adding libraries of standardised parts for specific model organisms [Lee et al., 2015; Moore et al., 2016; Rajkumar et al., 2019].

1.5 The Aim of this Study

Cells respond to environmental changes by adjusting their gene expression. One way cells respond is by regulating their transcription. It has been shown that cells respond to environmental changes by alterations in DNA supercoiling [Dorman, 1996]. Furthermore, research has demonstrated that changes in DNA supercoiling cause adjustments in gene expression [Sobetzko et al., 2013]. Moreover, the process of transcription itself produces DNA supercoiling on a local scale, which has been shown to influence the transcription of other genes in close proximity. These findings raise the question of whether cells actively use DNA supercoiling (globally and locally) for gene regulation and, if so, how they do this on the transcriptional level of gene expression. Moreover, if the mechanisms of regulation by DNA supercoiling were

able to be understood, it could open up new research avenues in the synthetic biology community. Finally, the question of whether supercoiling can be used for the regulation of genetic circuits or whether orthogonal genetic systems could be designed in such a way that they become less prone to unwanted effects caused by DNA supercoiling remains.

The main objective of this study was to establish an experimental pipeline that can investigate the effects of transcription-coupled and global DNA supercoiling on transcription at the transcription unit and the promoter level. In the best-case scenario, this experimental pipeline would both add to the knowledge of what makes promoters prone to regulation by DNA supercoiling and allow the production of promoters or genetic circuit setups that are controllable or immune to DNA supercoiling.
Firstly, a solution to issues of non-standardised cloning systems had to be created. This cloning system should be highly standardised but flexible enough to enable different arrangements of genetic parts to investigate expression behaviour in different TCDS contexts. The cloning solution needed to be compatible with existing standards following the synthetic biology idea of modularisation and standardisation.
Secondly, an experimental setup that allows for the systematic investigation of promoter parts and their influence on supercoiling sensitivity of promoters needed to be built. A *bona fide* target to focus on is the spacer part of the promoter. The setup should rule out as many as possible influences; therefore, it should use a minimal promoter embedded in a standardised genetic environment.
Lastly, a convenient way of integrating the arrangements into the genome should be implemented into the experimental pipeline. Preferentially this integration should be modular in a way that the arrangements built with the experimental pipeline described above could be easily transferred into the genome at variable positions to investigate differences between plasmid and specific chromosomal contexts.

2 Results

2.1 A Toolbox for Flexible and Standardised Modular Cloning

Parts of this section are published in [Klein et al., 2019]. For this publication I either did the experiments and cloning or supervised them when they were carried out by the bachelor students Leonie Emde or Aaron Kuijpers. I wrote the manuscript together with Dr. Patrick Sobetzko.

When I began my PhD, a modular cloning system to systematically investigate the influence of transcription onto neighbouring transcription was not available. The MoClo toolbox from [Weber et al., 2011], e.g. is appropriate for studying specific parts of transcription units like different ribosome binding site (RBS) or promoters, but is limited for investigations of effects on gene expression due to the arrangement of more than one transcription unit (see fig. 2.1). In the majority of the existing modular cloning systems, such as the standard MoClo system, the storage plasmid defines the position of a DNA fragment in the next level. The basic parts are stored in level 0 plasmids which can be combined and have a fixed destination stored in level 1 plasmids. Level 1 plasmids, usually storing transcription units, can then be assembled in the same way as level 0 plasmids, to networks or chromosomes (level 2) [Weber et al., 2011; Engler et al., 2014; Schindler et al., 2016; Messerschmidt et al., 2016]. On the one hand, the hierarchical structure of MoClo is a straight forward and automatable process; however, it may also lead to inflexibility and require massive cloning effort, e.g. if transcription units are to be tested in different arrangements. Thus, a method that allows free combination of DNA units and compatibility with

the standard MoClo system represented a methodological gap that we aimed to fill.

Hence, to overcome the innate inflexibility of the standard MoClo system, Flexible Modular Cloning (MoCloFlex) was build which added flexibility to Modular Cloning and also allowed the construction of plasmids *de novo* in a modular manner. Therefore, a set of new plasmids was introduced. The plasmid set comprised five position-vectors (MCF-positions) which were able to store any DNA fragment, 60 pre-built linker plasmids which allowed combination of two to five position vectors in any orientation and order, and a destination vector which allowed the storage of position assemblies. Thereby, the use of linkers is significantly reducing cloning effort (fig. 2.1). The combination of position vectors can form a new plasmid or be built into a destination vector. Our destination vector allows iterative rounds of cloning. Iterative cloning needs an extra cloning step from the destination into any position vector which can then be combined with other positions in a new destination. By the cost of an extra cloning step, the flexibility of the system is conserved into the following iterations, and no additional linkers or plasmids are needed. In the case that after some iterations, the flexibility of MoCloFlex is not needed any more, the construct stored in the destination vector can be transferred into any level 1 plasmids of the MoClo system and used in the system described by Weber et al.

During the construction of MoCloFlex, it was essential not to create another defined cloning standard, since there is a growing number of competing, incompatible cloning standards. Thus, from the onset compatibility with the existing MoClo standard was implemented.

2.1.1 The MoCloFlex Plasmids and their Application

The idea of MoCloFlex is to maximise flexibility when building gene arrangements, and at the same time, to allow standardization of the parts in such a way they are able to be integrated into the existing standard MoClo protocol. To achieve this, MoCloFlex comprises four plasmid classes: (1) MCF-Positions: can store DNA fragments and are either obtained and modified by Polymerase Chain Reaction (PCR) or through an

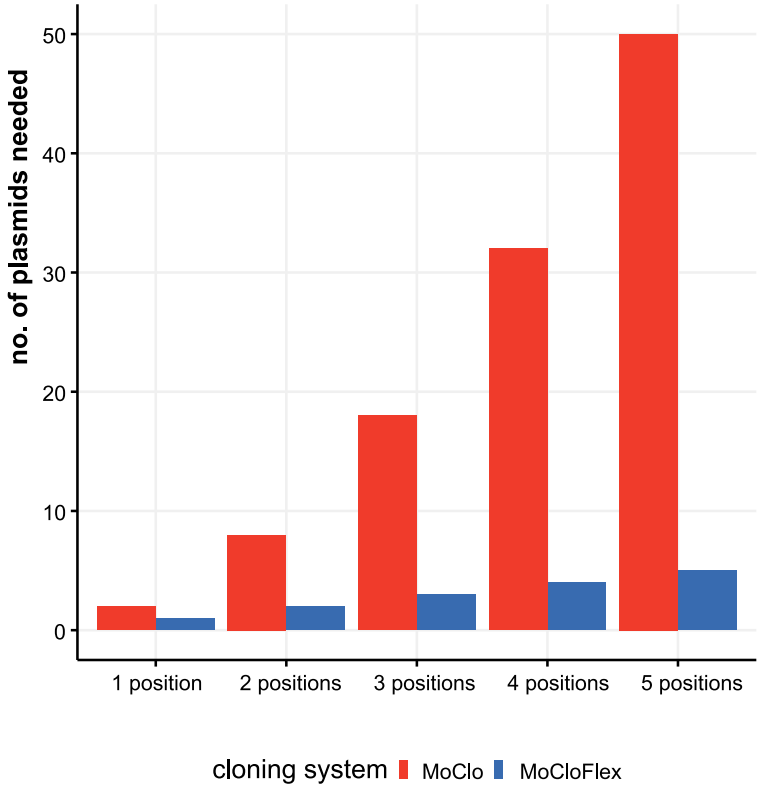


Figure 2.1: Construction effort needed to test combinatorial assemblies ranging from one to five fragments with standard MoClo compared to MoCloFlex. Testing genetic parts systematically requires investigation of every part at every position and in each orientation. Using MoClo, the number of plasmids required for such systematic investigation is $n = parts * positions * orientations$. In comparison, MoCloFlex requires $n = positions$ plasmids. With MoClo (red bars), the number of plasmids grows to up to 50 plasmids for testing five positions, whereas MoCloFlex (blue bars) only needs 5 plasmids for the same experiment.

existing MoClo Level 0 library. (2) MCF-Linkers: connect the different MCF-Positions to higher-order arrangements and mediate flexibility. (3) MCF-End-Linkers: allow building arrangements of MCF-Positions into the (4) MCF-Destination vector (fig. 2.2 a). There are two steps to build a DNA fragment into the MCF-Position. Firstly, a PCR is required to add a BsaI cut site and the entry motifs (GGAG and CTCG) for the fragment. Whatever fragment shall be built into an MCF-Position has to be free of BsaI, and BbsI recognition sites. The entry motifs are the same as those in level one of the standard MoClo system, which realise compatibility between MoClo and MoCloFlex. Thus an existing level 0 part collection form MoClo can be used to build into MCF-Positions since they share the same flanking motifs. Secondly, the PCR-fragment and the MCF-Position can be assembled by one-pot restriction-ligation (fig. 2.2 b). Note that it takes 24 h to obtain the final plasmid since the additional step of sequencing can be omitted due to the fact that no further PCR is carried out between the construction of the parts and the final assembly. Together, restriction-ligation ($5:00\ h$), transformation ($1:30\ h$), plating and incubation (overnight, $8:00\ h$), colony PCR ($1:45\ h$), growing right clones in liquid culture ($6:00\ h$), and plasmid prep ($1:45\ h$) could be done in one day.

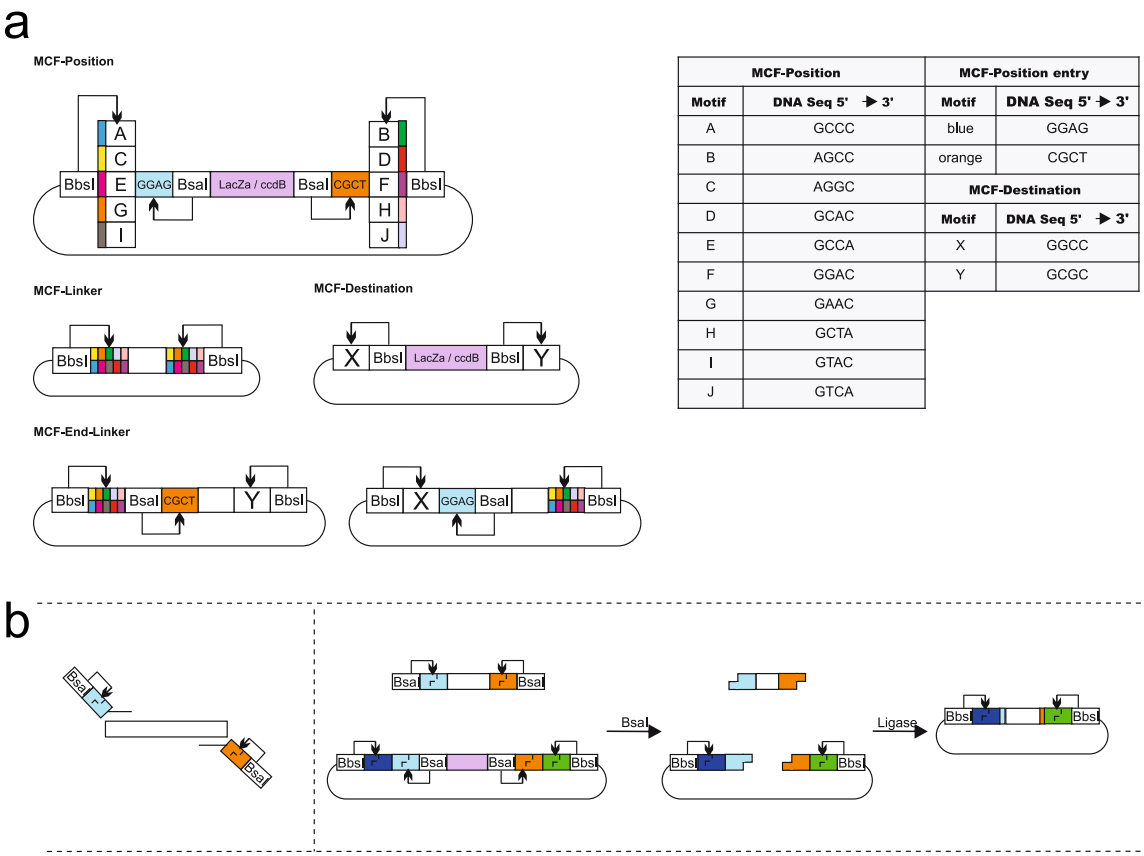
MoCloFlex has Two Modes of Action: Flexible Modular Cloning and Plasmid Assembly

MCF-Positions and MCF-Linkers can be used to build plasmids by storing all parts of the plasmid on different MCF-Position vectors and connecting them with MCF-Linker (fig. 2.4 a). To build e.g. networks or circuits of transcription units into the MCF-Destination vector, MCF-Linker and MCF-End-Linkers connect MCF-positions into the MCF-Destination vector. Thereby, the entry motifs for MCF-Positions are added at the borders of the construct, allowing rounds of iterative cloning. 40 MCF-Linkers and 20 MCF-End-Linker allows combination of any MCF-Position motifs and with the MCF-Destination vector. In appendix figure 4.4, the process is explained with an example guiding through the planing of a plasmid and the selection of MCF-parts for the one-pot reaction. The system allows incorporation of up to five MCF-Positions into the MCF-Destination vector at the same time (fig. 2.3 a). To avoid errors by mixing the systems, we chose only motifs for the restriction overhangs that are not part of the

MoClo system,with the exception of the entry motifs into the MCF-Positions, which are necessary to mediate compatibility. As another feature, the overhangs are prefix and suffix-free concerning BsaI and BbsI (BpiI) recognition sites, which excludes any accidental formation of new recognition sites and therefore assembly failures. (fig. 2.2a table).

2.1.2 Efficiency of Cloning MCF-Positions into the MCF-Destination Vector

We tested the efficiency of the restriction-ligation reaction with between a minimum of four DNA fragments and a maximum of 12 DNA fragments in one reaction. We built three MCF-Positions each with a different fluorescence protein regulated by an *aldA* promoter and two "dummy" positions containing sequences with no biological function. To test the cloning efficiency, one to five MCF-Positions were assembled into the MCF-Destination vector and the fraction of positive clones was determined. For the first construct 1 MCF-Position, we combined 4 DNA fragments in the restriction-ligation: The MCF-Destination, two MCF-End-Linkers (XA and BY) and the MCF-Position AB. Each additional MCF-Position requires another MCF-Linker; thus, two additional MCF-Linker vectors must be added to the reaction for every additional MCF-Position in the restriction-ligation. We calculated the cloning efficiency from clones that displayed the expected fluorescence as a fraction of total colonies on the plates. For testing the cloning efficiency of 10 and 12 DNA fragments per reaction also sequencing was used to confirm the results. For 4 DNA fragments (1 MCF-Position) the efficiency was 92.3 $\pm$ 3 %; for 6 DNA fragments (2 MCF-Positions) in the restriction-ligation the efficiency was 92.6 $\pm$ 2.3 %; for 8 DNA fragments (3 MCF-Positions) the efficiency dropped to 51.2 $\pm$ 8.1 % and and further dropped to 33.7 $\pm$ 4.5 % when 10 DNA fragments (4 MCF-Positions) were used. The maximum construct built in our system in one reaction was 12 DNA fragments (5 MCF-Positions), with an efficiency of 8.8 $\pm$ 6.2 % (fig. 2.3b). Notably, the number of colonies also dropped from 2021 colonies with 4 fragments to 189 with 12 fragments per reaction. Since the constructs we chose contained highly repetitive sequences such as mTurquiose2, mVenus, mCherry, identical promoters, and terminators, which make a construct



MCF-Position		MCF-Position entry	
Motif	**DNA Seq 5' → 3'**	**Motif**	**DNA Seq 5' → 3'**
A	GCCC	blue	GGAG
B	AGCC	orange	CGCT
C	AGGC	**MCF-Destination**	
D	GCAC	**Motif**	**DNA Seq 5' → 3'**
E	GCCA	X	GGCC
F	GGAC	Y	GCGC
G	GAAC		
H	GCTA		
I	GTAC		
J	GTCA		

Figure 2.2: (a) Overview of MoCloFlex: MCF-Positions contain an inner insertion cassette accessible through insertion motifs (blue and orange; BsaI). The insertion cassette is flanked by combination motifs (A - J; BbsI). Every MCF-Linker has one combination of two motifs A - J. MCF-End-Linkers contain one destination motif (X or Y; BbsI) for any combination motif (A - J) and have one integration motif. MCF-Destination plasmids have a *ccdB* and *lacZ* gene for selection and counter-selection flanked by the destination motifs X and Y (BbsI). (b) Any DNA fragment to be inserted into an MCF-Position needs to get the insertion motifs by PCR. When cut with BsaI and ligated with one another, both MCF-Position and fragment lose their BsaI recognition sites, and the fragment replaces the integration box of the MCF-Position.

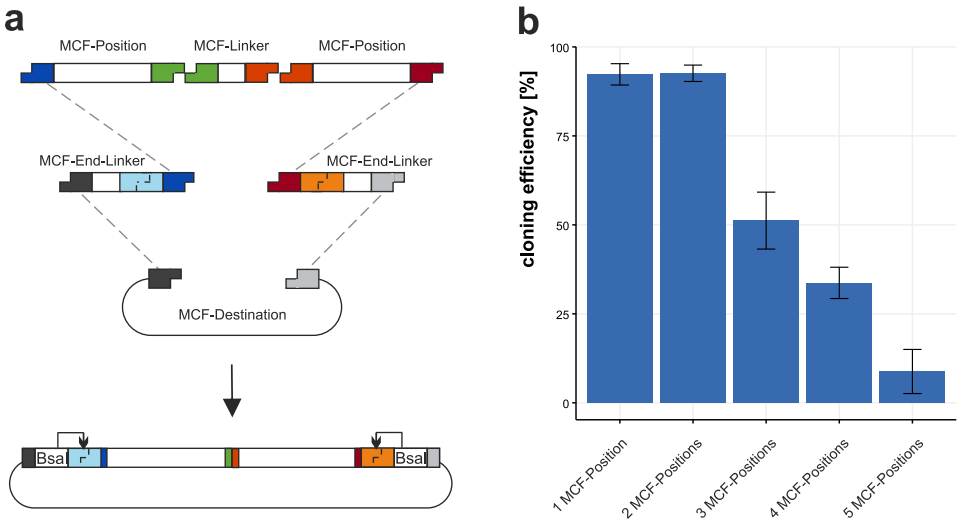


Figure 2.3: Use of the MCF-Destination to build arrangements of MCF-Positions. (a) To combine MCF-Positions into the MCF-Destination, MCF-End-Linkers are fusing MCF-Positions with the MCF-Destination vector and the MCF-Linker is fusing the MCF-Positions. (b) The cloning efficiency of 1 to 5 MCF-Positions (4-12 DNA fragments) into the MCF-Destination vector. Efficiency is defined as % of positive clones per tested clones. Error bars represent standard deviation from 3 individual experiments with 95 clones each.

difficult to clone, cloning efficiency is likely to be higher when using MoCloFlex with non-repetitive sequences.

2.1.3 Cloning Efficiency of *de novo* Plasmid Assembly

To test the efficiency of building plasmids with MoCloFlex, using only MCF-Positions and MCF-Linkers, we created a plasmid out of 6 MCF-Parts: MCF-Position AB which contained an origin of replication (p15A), MCF-Position CD which contained a chloramphenicol resistance cassette, MCF-Position EF which contained a CDS of mCherry, and 3 MCF-Linkers (BC, DE, FA) to bridge between the MCF-Positions. When we first built plasmids using this setup, background colonies that carrying the uncut MCF-Position with the chloramphenicol resistance cassette appeared, which reduced the efficiency from 50 % to around 35 % (fig. 2.4 b). To solve this issue, we integrated an I-SceI recognition site [Monteilhet et al., 1990] into the MCF-Position CD and

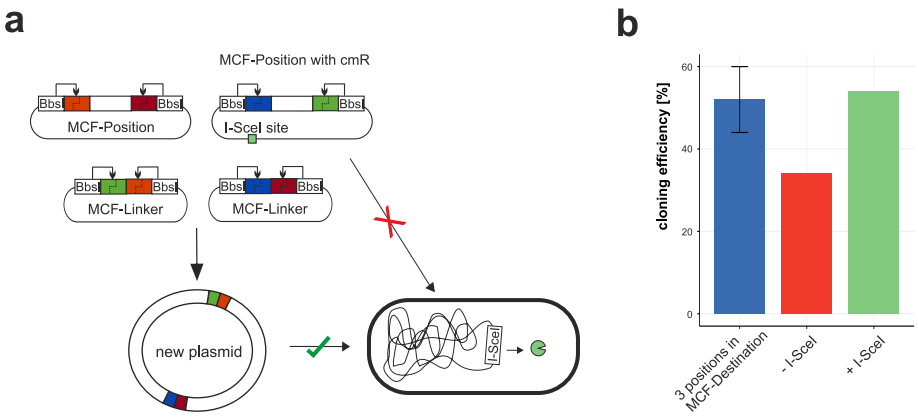


Figure 2.4: Combining MCF-Positions and MCF-Linkers to a new plasmid. (a) The plasmid carrying the antibiotic resistance marker cmR cannot be maintained in a strain expressing I-SceI, thus reducing background colonies and restoring cloning efficiency (b).

transformed our constructs into an *E. coli* strain which expressed I-SceI meganuclease [Monteilhet et al., 1990]; this avoided cloning background with uncut MCF-Position plasmids (fig. 2.4 b) and restored efficiency to the 50 % expected when using 6 fragments in one reaction.

2.1.4 RecBCD Digestion Increases Cloning Efficiency but Decreases Number of Clones

The sequencing of negative clones revealed recombination events which could not be explained by the *in vitro* restriction-ligation reaction. Hence, the recombination probably occurred *in vivo*. To test whether partially assembled linear fragments recombined *in vivo* after transformation, we incubated the reaction after the restriction-ligation with RecBCD. RecBCD is an enzyme complex that in *E. coli* partakes in homologous recombination but also has a nuclease function for double-strand and single-strand DNA [Yu et al., 1998; Amundsen et al., 1986]. RecBCD incubation previous to transformation of a 8 DNA fragment assembly (3 MCF-Positions) decreased the number of colonies from around 250 to 10 but also increased the cloning efficiency dramatically from 50 % to 90 %. From these preliminary results, we inferred that ligation or recombination of linear DNA fragments occurs to a significant level after transformation. The capability of *E. coli* to make plasmids out of double stranded

DNA is exploited by various *in vivo* cloning methods, e.g. [Beyer et al., 2015] and the mechanism is shown to be dependent on the exonuclease III XthA [Nozaki and Niki, 2019] which is present in our cloning strain. Hence, we hypothesised that the unwanted recombination observed was due to this mechanism, as it is dependent on sequences which can be found in the fluorescent proteins, the promoters, and the terminators used in this study. Nevertheless, following use of RecBCD, colonies only appeared when a maximum of 8 fragments in the restriction-ligation was used. Thus RecBCD digestion is not applicable for improving the cloning efficiency of more than 3 MCF-Positions in one reaction.

2.1.5 The Arrangement of Transcription Units in a Network Influences their Expression

To test the flexibility of MoCloFlex, we decided to build three networks of three different transcription units in three arrangements. The first transcription unit consists of *gyrBp* controlling the expression of mVenus and a T0 terminator. The other two transcription units were either mCherry or mTurquoise2 controlled by *aldAp* and a T0 terminator. Both promoters appeared in the [Blot et al., 2006] data set as upregulated by relaxation. The promoter sequence was chosen from the promoter annotation found in the database `www.ecocyc.org` but cloned into the constructs without their native RBS. In the first construct, both flanking transcription units pointed to the *gyrBp* controlled transcription unit, which is called a convergent arrangement. Second, we arranged both *aldAp* controlled transcription units in a divergent orientation with respect to the *gyrBp* cassette. In the last orientation, all cassettes pointed in the same direction, which is called tandem orientation (fig. 2.5 a). As shown in fig. 2.5 b, we observed differential expression patterns in all three arrangements with a maximal expression for all three transcription units when in tandem orientation. The arrangements were maintained on plasmids containing the P15A origin of replication, which is closely related to the ColE1 origin of replication [Selzer et al., 1983]. The P15A origin leads to around 10 copies of the plasmid per cell. Read-through transcription can interfere with the replication initiation in ColE1 [Stueber and Bujard, 1982]. Inferring from this, it could be that different transcription arrangements lead to slightly different copy numbers, thus changing the expression level. However, our transcription units contained terminators

protecting the origin of replication in every tested arrangement, and they did not allow for 'read-through' transcription as demonstrated in [Emde, 2017]. Another possibility may be that the plasmid topology could be altered by the arrangement of three transcription units, since almost half of the *E. coli* promoters, including the *gyrA* promoter, respond to altered DNA supercoiling either induced globally or by neighbouring expression [Dages et al., 2018; Lim et al., 2003; Sobetzko, 2016]. Alternatively, it could be that the differences in expression levels are due to *aldA* and *gyrB* promoters respond to different DNA supercoiling levels they get from their neighbours, but this should be investigated further in a chromosomal context. However, there was a measurable alteration in the expression levels of our three constructs, and this was dependent on the arrangement of the transcription units, which confirmed the need for testing arrangements systematically. Hence, arrangement and orientation matters and could be easily be screened with MoCloFlex.

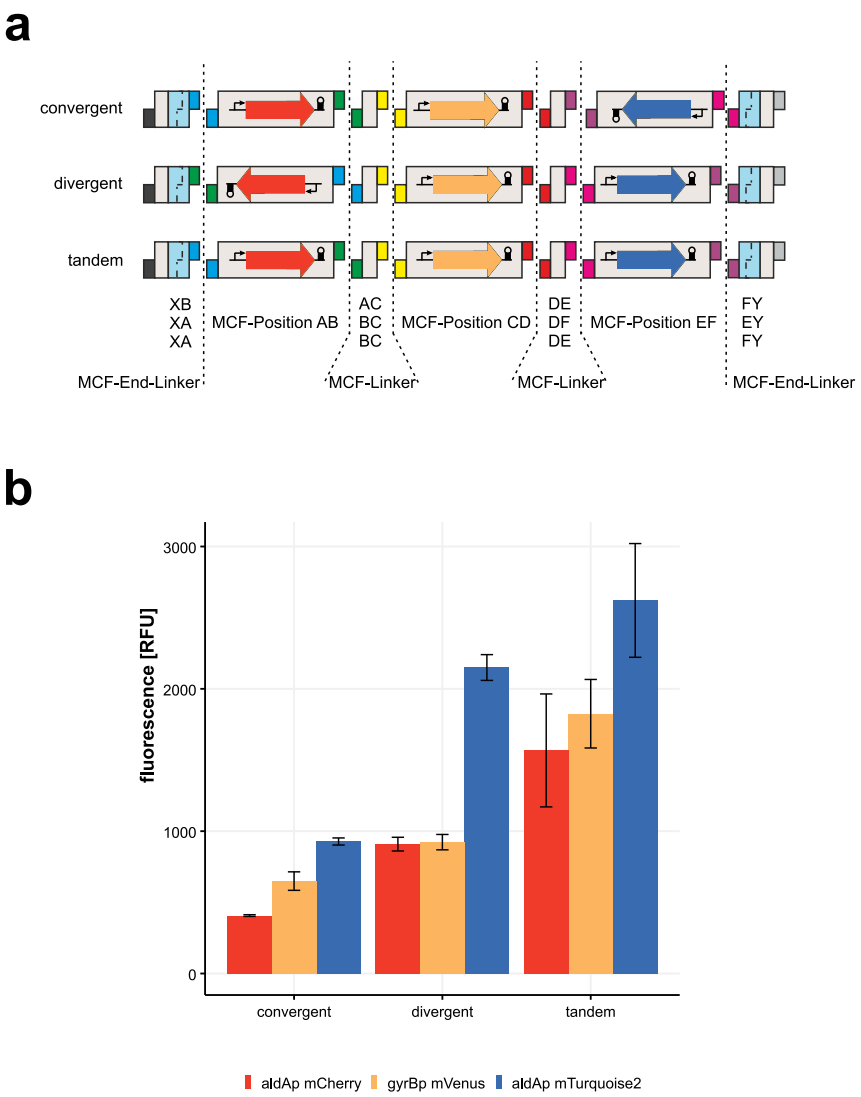


Figure 2.5: Expression of three transcription units in three different contexts. (a) list of MCF-Linkers, MCF-End-Linkers and MCF-Positions, which were built into the MCF-Destination vector. Convergent: *aldAp* controlling mCherry and *aldAp* controlling mTurquoise2 expression, pointing towards *gyrBp* controlling mVenus expression. Divergent: *aldAp* promoters pointing away from *gyrBp*. Tandem: every transcription unit points in the same direction. (b) Expression in (RFU) of the three transcription units in the three different arrangements.

2.2 What Makes a Promoter Respond to DNA Supercoiling?

For more than half of the *E. coli* genes, it has been shown that they react to DNA-supercoiling changes by up- or down-regulation of their expression [Blot et al., 2006]. For some individual promoters it has also been found, that they respond to neighbouring expression, possibly by TCDS [Dages et al., 2018]. However, the promoter parts which contribute or have the most significant influence has not been simple to determine. On a bacterial chromosome, each gene or operon can often be under the control of more than one promoter, depending on the individual context. Also the individual promoters differ in their sequence of motifs. One *bona fide* target is the spacer region since it arranges the -35 and the -10 parts of the promoter. For single promoters, it has been shown that changing the length simply by adding or deleting a single nucleotide (nt) altered their supercoiling sensitivity [Aoyama and Takanami, 1988]. However, a more systematic approach to investigate promoter parts and their effects on supercoiling-sensitivity is lacking.

To tackle the question the pNovoScreen-plasmid was built, which was first characterised and then used to investigate the spacer region of a minimal promoter in a systematic approach.

2.2.1 Construction and Characterization of the NovoScreen-Plasmid

To begin with, the pNovoScreen plasmid was built. The backbone was constructed by fusing a P15A origin of replication (ori) next to a chloramphenicol resistance cassette in tandem orientation. All parts which were ordered as oligonucleotides were freed of type IIs restriction sites. Whenever such a restriction site was found in an open reading frame (ORF) or part amplifyed by PCR, a silent mutation consisting of a substitute codon with similar tRNA frequency was inserted by mutagenesis PCR. Two fluorescent reporter genes (mVenus and mCherry) were built in convergent orientation and separated by rho-independent terminators. A *colE* promoter controlled

the mCherry fluorescent protein (fig. 2.6B). The synthetic spacer landing pad was built in front of the mVenus gene. The cloning site displayed in figure 2.6A was constructed using oligonucleotides. For the -35 and the -10-region, the consensus sequences (-35: 5'-TTGACA-3', -10: 5'-TATAAT-3') were used. This minimal promoter was then flanked with the restriction sites of the MoClo level 0 plasmid pICH41233 (BbsI; GGAG, TACT) [Weber et al., 2011]. These cut sides allowed any promoter from this screen to also be used as a part of a MoClo level 0 promoter collection. The inner restriction sites were required for insertion of the spacer libraries as depicted in figure 2.6A. The *LacZ*α is removed by insertion of a library, thus making positive clones appear white when isopropyl β-d-1-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) are supplemented to the agar (Blue-White Screening). The libraries were ordered as oligonucleotides (Primers CK 173-175, CK267-CK277). To make them double-stranded DNA (dsDNA) they were used in a PCR with CK176. In a one-pot restriction ligation reaction the libraries were built into the plasmid and afterwards transformed in chemically competent TOP10 cells. Following transformation, colonies were picked and used to inoculate 100 µL LB in 96-well plates, which were stored as master-plates following incubation by adding 100 µL 50%(v/v) glycerol. From these master plates, the experimental plates were inoculated and fluorometric assays performed (see protocol 4.5).

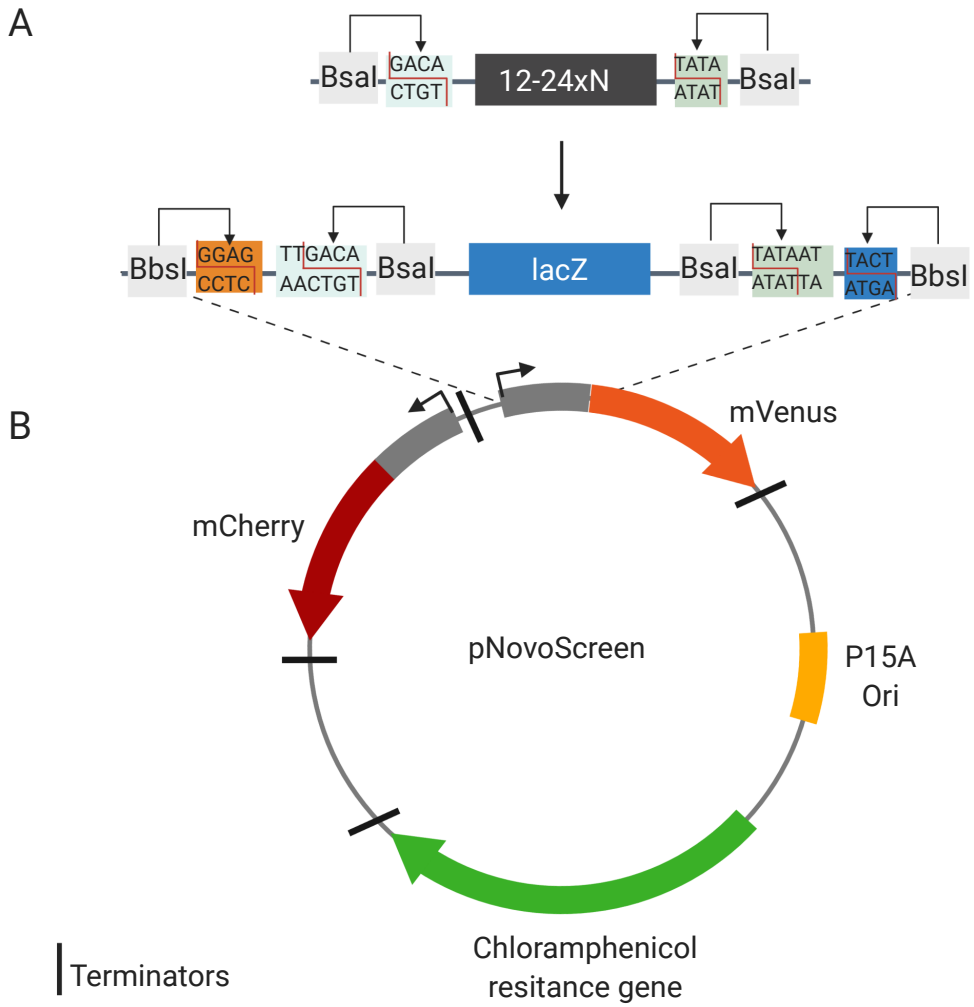


Figure 2.6: (A) Spacer library preparation for synthetic σ^{70}-promoters. For the different spacer lengths, oligonucleotides of random sequence and different length were built into the screening plasmid using a golden-gate reaction leaving no scar. The random fragment replaced the *LacZα* gene for convenient selection. (B) pNovoScreen plasmid map. mCherry reporter was under control of *ColEp* expression.

2.2.2 The Promoter Spacer Length and Sequence Changes the Expression Strength

Firstly, the influence of the spacer region (length and sequence) on promoter strength in our synthetic promoters was examined. To choose the spacer lengths we extracted all σ^{70}-promoters from regulonDB[1] [Salgado et al., 2018] and calculated the spacer lengths of endogenous promoters (predicted ones and experimentally proven). The distribution of spacer lengths from this data set is given in figure 2.7B. Endogenous promoters have a spacer length between 15 and 21 bps; this was extended to construct libraries containing spacers of lengths between 12 and 23 bps. A 150 bp random sequence was used instead of our minimal promoter as a control for background expression. As depicted in figure 2.7A, different spacer lengths and sequences led to substantial variations in mVenus expression. The sub-optimal spacer lengths of 12-14 bp and 22-23 bp showed low to no expression above the control; however, surprisingly the spacers 15 bp, 20 bp and 21 bp showed also shallow expression, which was unlike the expression patterns of endogenous promoters of the same spacer length (fig. 2.7C). The overall variations in expression were stronger in the synthetic promoters than in endogenous promoters of different spacer lengths. Most active promoters appeared in the spacer lengths of 17 bp and 18 bp, whereby the 18 bp spacers were on average a little stronger than the 17 bp spacers.
For comparison, the influence of spacer length on endogenous gene expression was investigated. A subset from all promoters in regulonDB was created by searching only σ^{70}-promoters and allowing three mismatches (hamming distance) to the consensus -35 and -10-sequences. In our analysis, it did not matter if the mismatches were in the -35, -10, or both since this made no clear difference. This subset of promoters was then merged with the transcriptomics expression data from [Sobetzko et al., 2013]. In the cases where there were more promoters for the same gene, the σ^{70}-promoter was chosen. The strongest promoters were among the 17 bp and 18 bp promoters. The expression of promoters with shorter spacers was lower and longer spacers are almost equally strong as the 18 bp spacer promoters.

[1] `http://regulondb.ccg.unam.mx/`

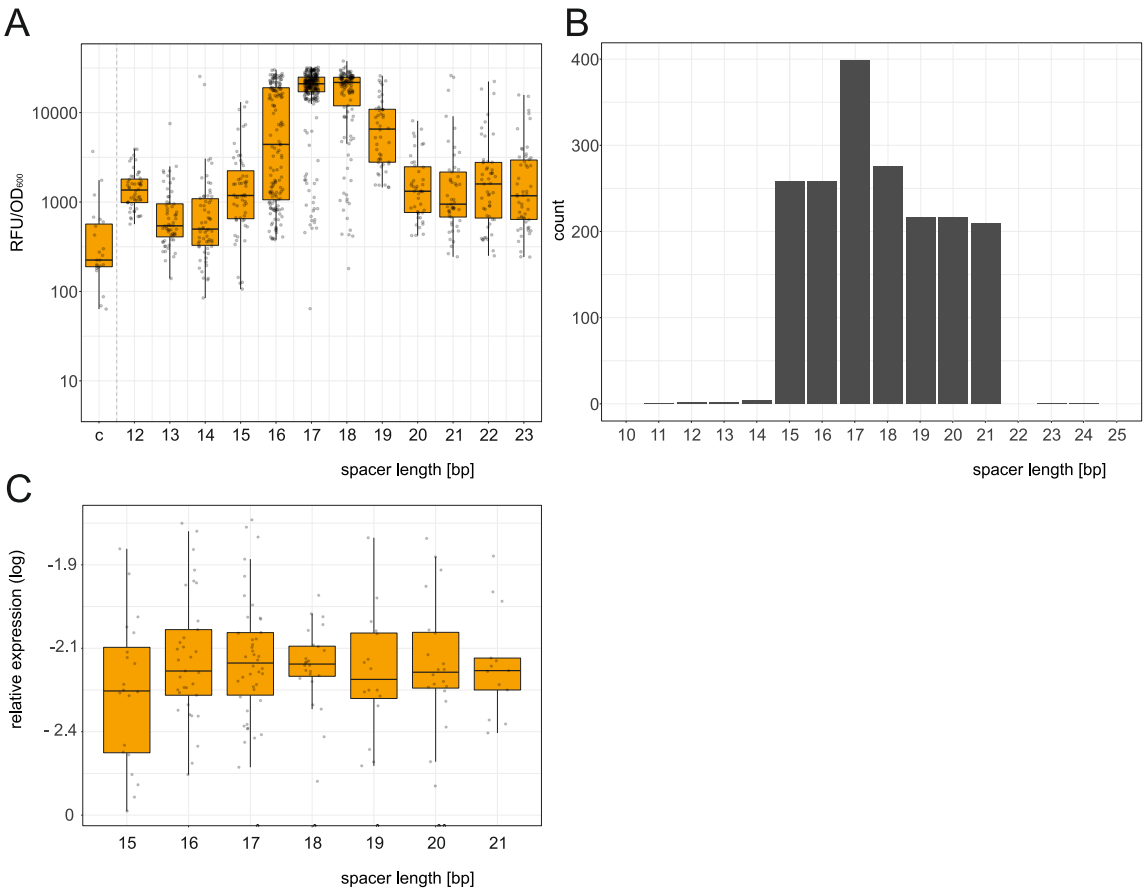


Figure 2.7: (A) The different spacer lengths influenced the expression strength of the mVenus fluorescence protein that is downstream of the synthetic promoters. The spacers with 16 to 19 bp length displayed the most active promoters.(B) Spacer length distribution from all σ^{70}-promoters in *E. coli*. Data was extracted from RegulonDB (C) Relative expression strength of endogenous σ^{70}-promoters with different spacer lengths. For better comparability, only promoters with a hamming distance of less than 3 in their -10 and -35- sequences were considered from the data out of [Sobetzko et al., 2013]. The expression data from [Sobetzko et al., 2013] is normalised that expression of all genes in this study is one. Since the analysis here, used a subset of these promoters the expression together is less than one.

2.2.3 The Spacer has Only a Small Impact on Relaxation Sensitivity

The spacing between both σ binding motifs is said to influence the supercoiling sensitivity of a promoter which was shown for single promoters. However, it is unknown whether this is due only to the spacing, or whether the sequence of these spacers, e.g. motifs or certain bases, may play a role, as older studies focused on a limited set of promoters and did not examine the sequence systematically [Aoyama and Takanami, 1988]. To test whether the spacer was mediating supercoiling sensitivity in this screen, previously built promoter libraries were studied under DNA-relaxing conditions with the drug novobiocin. Promoter libraries were mirrored onto two 96-well plates and added a sublethal concentration (17 μg/mL) of novobiocin to one of the duplicates. Novobiocin inhibits the function of gyrase and thus relaxes DNA in the cell. Then, cells were let grown overnight for 15 h (see methods 4.5.1). At the chosen concentration, no growth defect could be detected during a 15 h growth [Emde, 2017]. Following this, OD and fluorescence were measured in both plates, measurements were normalised to OD and ratios of expression were calculated using script NovoScreenv1.R (see 4.12.1). Plotting these ratios for the different spacer lengths revealed that in almost all spacer lengths, the mVenus expression decreased under DNA-relaxation between less than 1.4-fold to 2-fold (fig. 2.8A). The 18 bp spacers displayed the sharpest decline followed by 17 bp spacers and 19 bp. There was a tendency for promoters with the not optimal spacer lengths 12–14 bp and 20–23 bp sequences to show less reduction in expression when treated with novobiocin (fig. 2.8 A and B). Overall, there was a strong negative correlation between gene expression and relaxation sensitivity (spearman correlation coefficient -0.644, fig. 2.8A. This negative correlation was observed for individual spacers as well as for all spacers combined. When corrected for the effects of expression strength, there was a corridor for each spacer in which the supercoiling sensitivity altered expression up and down to 1.4-fold (fig. 2.8C and D). This change could be due to the sequence. Notably, even though the 17 bp and 18 bp spacer promoters are equally active, the 18 bp spacer showed a stronger drop in their expression before and after the data was corrected.

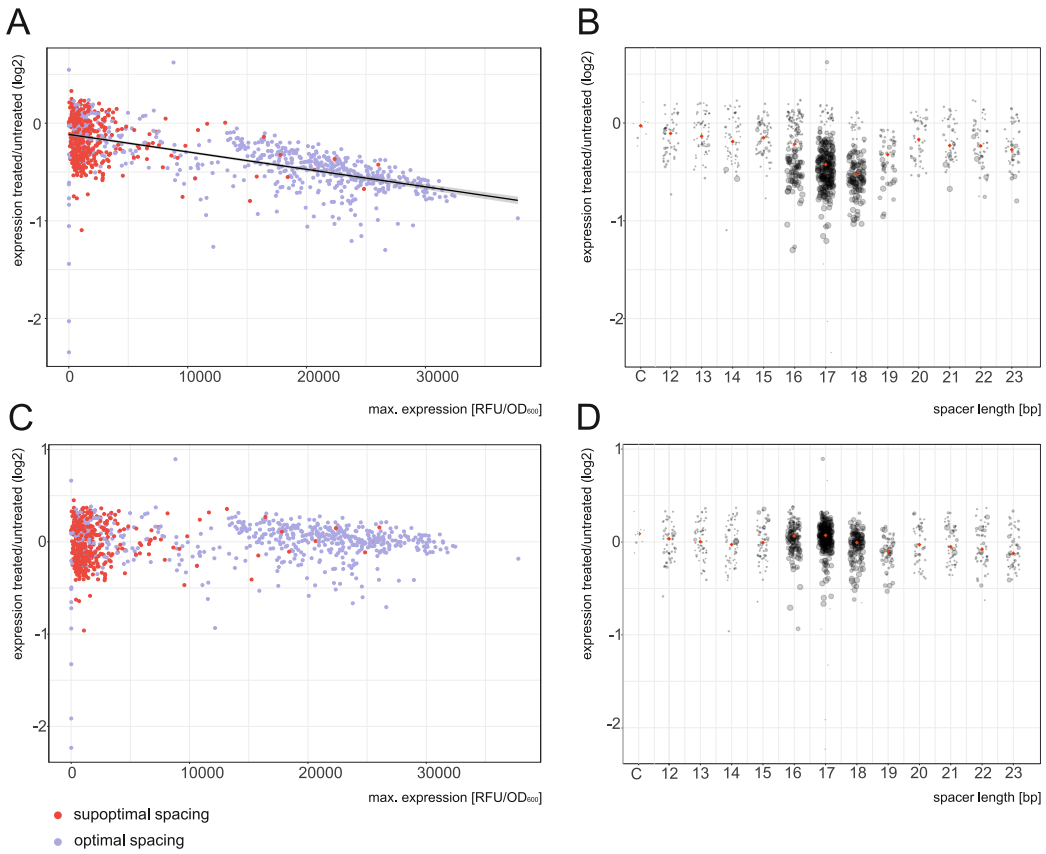


Figure 2.8: Supercoiling sensitivity of the synthetic σ^{70}-promoters is dependent on expression strength. In **(A)** promoters show more supercoiling sensitivity the stronger they are. The black line is the linear regression and shows a correlation between expression strength and supercoiling sensitivity (relaxation). In **(B)** the relationship of supercoiling sensitivity and expression strength is shown for spacer length library. The red dots indicate mean relaxation sensitivity and the circle size the expression strength (the bigger, the stronger). **(C)** The same plot as in **(A)** but corrected for the expression dependency. **(D)** The plot as in **(B)** but also adjusted for the expression dependency.

2.2.4 Sequencing of the 17 bp-Spacer Promoters Revealed a 5′-TGTG-3′-Motif

The 17 bp spacer is the most abundant endogenous spacer of all σ^{70} promoters in *E. coli* (fig. 2.7B). In our synthetic promoter library, the 17 bp spacer was on average equally strong or weaker than the 18 bp spacer promoters (fig. 2.7A). Compared to the endogenous data where the 17 bp spacer promoters were on average stronger than the 18 bp spacer (fig. 2.7C), the question was what may be causing this discrepancy. A new library of 17 bp promoters was constructed and expression strength and supercoiling sensitivity was tested. As in the expression test, the 17 bp spacer promoters again displayed a strength-dependent relaxation sensitivity (fig. 2.9 A). To then test how spacer sequence influences promoter strength and which sequence makes the weak 17 bp promoters weak and the strong ones strong, the library was divided into three groups: the least active 48, the most active 48, and all others. From the weak and the active group, a new master plate was made, and three technical replicates of the relaxation sensitivity measurements were performed. The results depicted in figure 2.9B show that there was some variation in expression and sensitivity, but still, the promoters could be distinguished from one another, which can only be explained by the different spacer sequences. These different spacers were then sent for sequencing, and the spacer sequences were extracted from the reads using R (script 4.12.2, sequences given in table 4.13). At first, we made web logos using the gglogo[2] package reviewing if any positions prefer distinctive bases. Surprisingly, both groups displayed a 5′-TG-3′ motif on location 15 and 16 of the spacer. This motif is found in so-called extended -10 promoter regions. The right group (strong promoters) contained this motif in 75 % of sequences, whereas 100 % of promoter sequences in the left group displayed this motif at spacer positions 15 + 16. Furthermore, the left group showed significant enrichment not only for one 5′-TG-3′ on this position but also for another 5′-TG-3′ just in front of the other one, leading to a 5′-TGTG-3′-motif. There were also promoters which contained more than two 5′-TG-3′ motifs in their sequence. No significant third position for these motifs in the spacer was found (fig.2.9 D). In both the left and right group combined, the 5′-TG-3′-motif was in more than 90% of the sequenced promoters on position 15 and 16. This raised the question of whether

[2] `https://CRAN.R-project.org/package=gglogo`

the oligonucleotides used to make the library were not randomised enough in this position. Therefore, we also tested the library in a no-selection context (not resulting in a promoter). Sequencing of 10 clones from this no-selection backbone revealed an almost equal distribution of all four bases for all positions of the 17 bp spacer (appendix fig. 4.3). Thus, the oligonucleotides could not explain this enrichment of 5'-TG-3'-motifs. An analysis of the melting, the bendability and the GC-content, revealed no significant difference between the sequences of the left compared to these of the right group (fig. 2.10 A.B, and C). Neither the mean values could distinguish between the two groups, nor a nucleotide resolution of bendability scores and melting energy revealed a clear distinction (fig. 2.10 E and F).

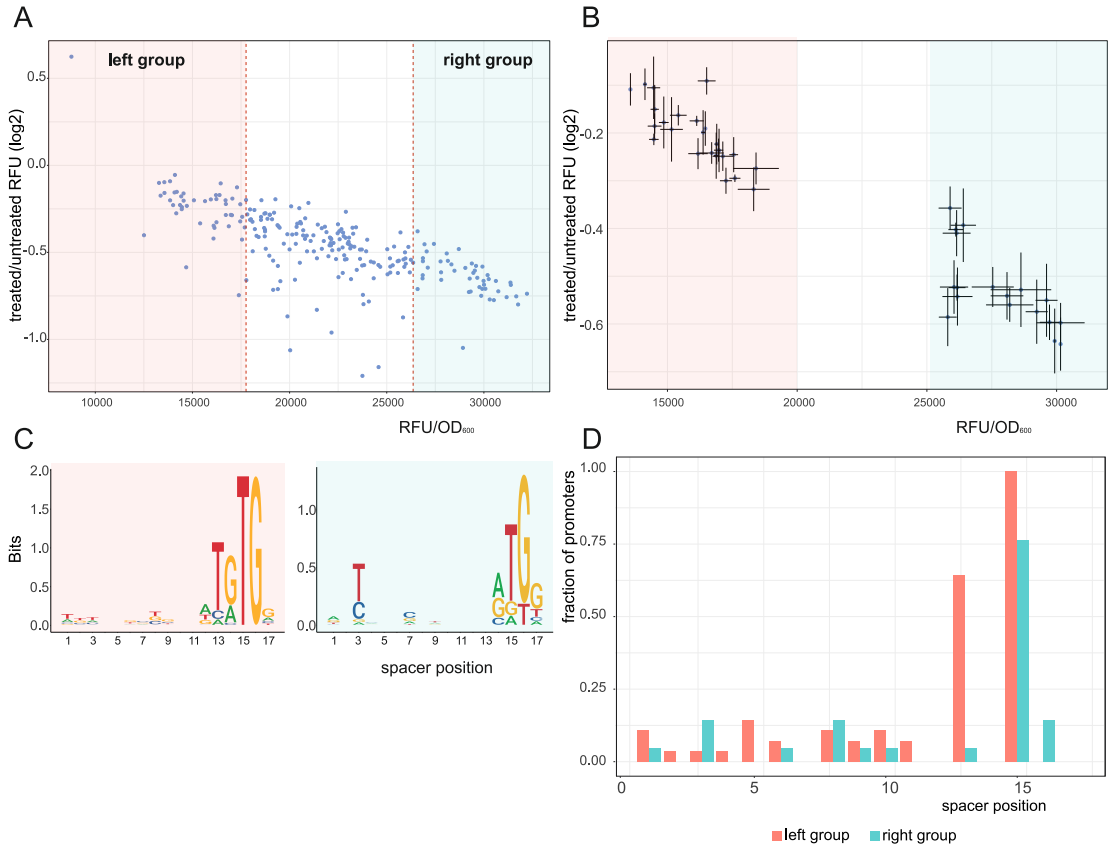


Figure 2.9: Sequence logo analysis revealed 5′-TG-3′- and 5′-TGTG-3′-motifs in the 17 bp spacer. **A** The relaxation-sensitivity and expression strength of 17 bp spacer library promoters. **B** Three replicates of the measurements from the left and right group defined in **A**; lines indicate standard error. **C** Sequence logos of the spacers from the confirmed promoters from **B**. **D** The distribution of 5′-TG-3′-motifs in the left and right group.

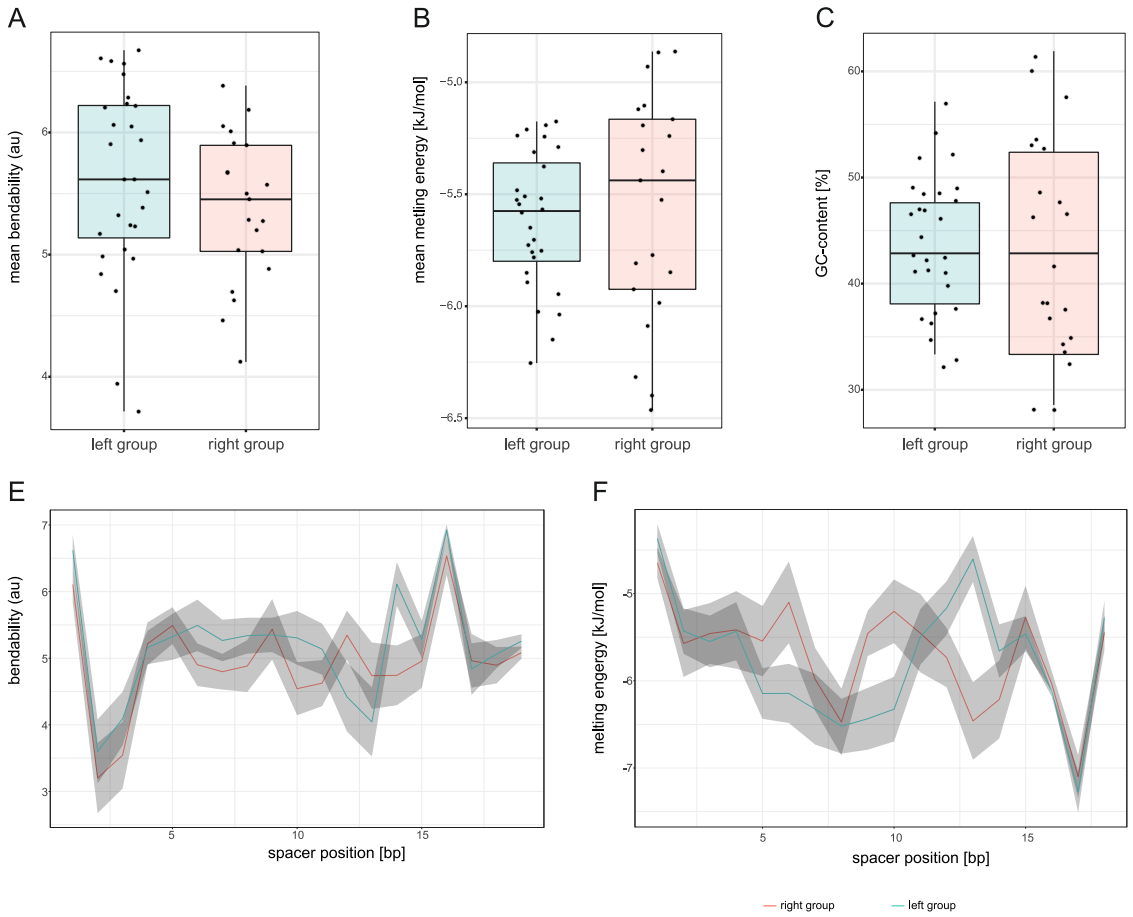


Figure 2.10: The mean bendability score (**A**), mean melting energy (**B**), and mean GC-content (**B**) of the 17 bp spacer library revealed no significant differences between left and right group sequences from figure 2.8. Melting energy (**E**) and bendability (**F**) as examined per nucleotide; both, the bendability score and the melting energy were calculated using a sliding window of three nucleotides. Values for nearest neighbour melting energy per dinucleotide at 37 °C were taken from [SantaLucia, 1998] and for bendability from [Gabrielian et al., 1997, 1996]. Grey area represents standard error.

2.2.5 High Transcription can Lead to Translation-Independent mRNA-Toxicity

Inferring from the cutoff in the 17 bp spacer expression data and the slightly fewer colonies observed after the 17 bp library transformations, it seemed that our promoters reached a maximum expression, with further expression being lethal. This may be due to the resulting levels of protein from the promoter expression, since overexpression of any protein can be toxic due to its resource costs for the cell [Stoebel et al., 2008]; for GFP, this limit is said to be 15 % of all proteins [Bolognesi and Lehner, 2018]. Since no knowledge about how much mVenus protein in % of total protein in the cell was produced by the most active promoters and if mVenus toxicity is similar to that of the described GFP, it was tried to lower the translation rate by changing the RBS in the pNovoScreen plasmid to one resulting in fewer translation events [Sobetzko, unpublished]. With the pNovoScreen-RBS7 plasmid, we repeated the screen for the 17 bp spacers and took the strongest sequenced promoter and the weakest promoter from the previous screen as reference promoters. It was hypothesised that if the translation were the problem, the weak promoter reference would remain one of the weaker promoters,but that other promoters would lead to higher expression than the strongest promoter from the previous screen. Surprisingly, we saw that both reference promoters kept their relative positions within new promoter libraries (fig. 2.11). Thus, we hypothesised that mRNA levels may have been the limiting factor. Notably, one recent publication has linked fluorescent protein GFP with mRNA toxicity [Mittal et al., 2018]. If their findings were also to apply to mVenus and what part of the mRNA exactly causes the toxicity, would be an interesting question to follow. However, we accidentally implemented a limiting factor in our screen that, at least for the 17 bp spacers, allows only a maximum of mRNA levels produced by the promoters. Moreover, the 5'-TGTG-3'-motif found by sequencing seemed to be a consequence of this selection pressure.

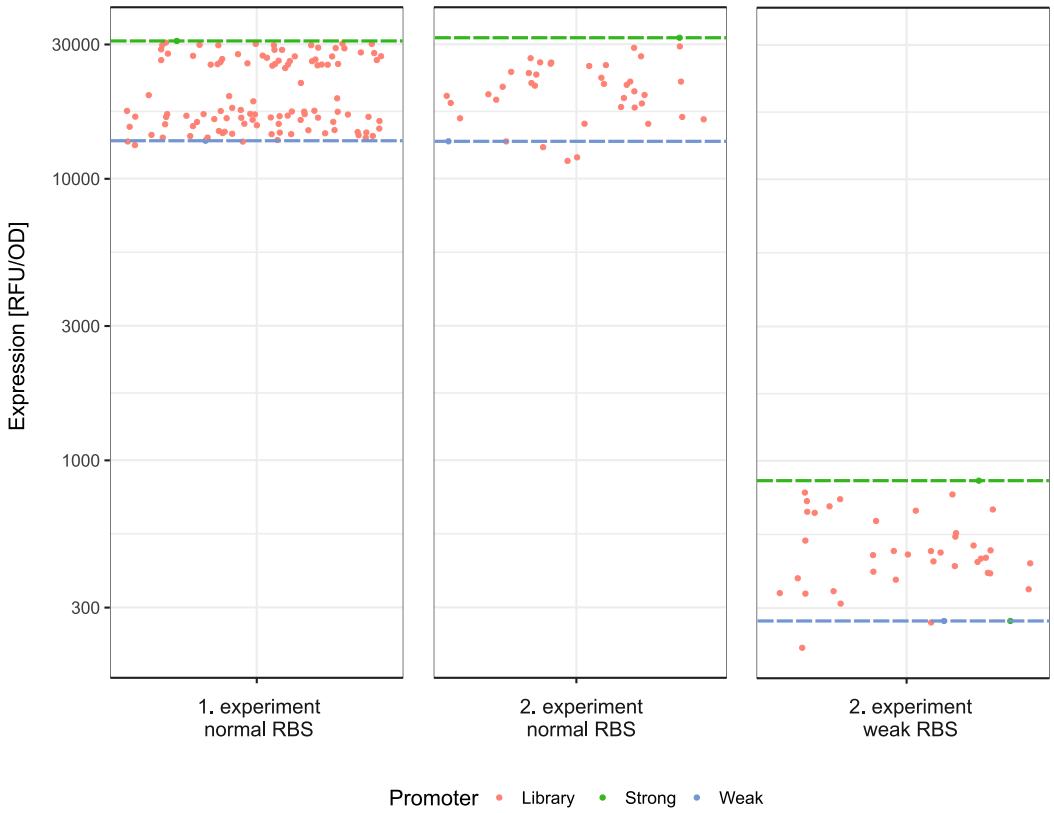


Figure 2.11: Lowering translation by changing the RBS did not change the relative expression of the reference promoters compared to a random library. The most active 17 bp spacer promoter remained the strongest when tested against a library in the context of fewer total translation events. The strongest and one of the weakest promoter from the 17 bp spacer library were used (left panel) as references when changing the RBS to a weaker one (right panel). As a control, a new library was built into the old backbone, and the reference promoters were again cloned into the old backbone (middle panel).

2.2.6 Promoters with 5′-TGTG-3′-Motifs can Act as RNAP Brakes

It was hypothesised that the enrichment of 5′-TG-3′- and 5′-TGTG-3′-motifs in the sequences of the 17 bp spacers, lower the transcription of genes controlled by these promoters. Since the 5′-TG-3′-motif interacts with the sigma factor, it could be that 5′-TG-3′ together with the consensus sequences of the -10 and -35 region result in a strong interaction that entraps the RNAP on the promoter. Following this logic, more TGs could lead to even stronger interactions and a more effective entrapment of the RNAP on the promoter. This entrapment would, as a consequence, lead to lower transcription rates and less expression. To test this hypothesis, we performed EMSA (methods section 4.3.1) using commercial RNAP enriched with σ^{70} from NEB and the strongest and the weakest promoter from our 17 bp spacer promoter library. Notably, the weakest promoter contained a 5′-TGTGTG-3′ triplet between spacer position 11 to 16, and the strongest contained one 5′-TG-3′ at positions 15 and 16. Testing both sequences by EMSA with RNAP revealed a stronger affinity to the enzyme for the weaker promoter (KD = 100 nM) by a factor of 7 compared to the KD of the most active promoter (KD = 700 nM) (fig. 2.12).

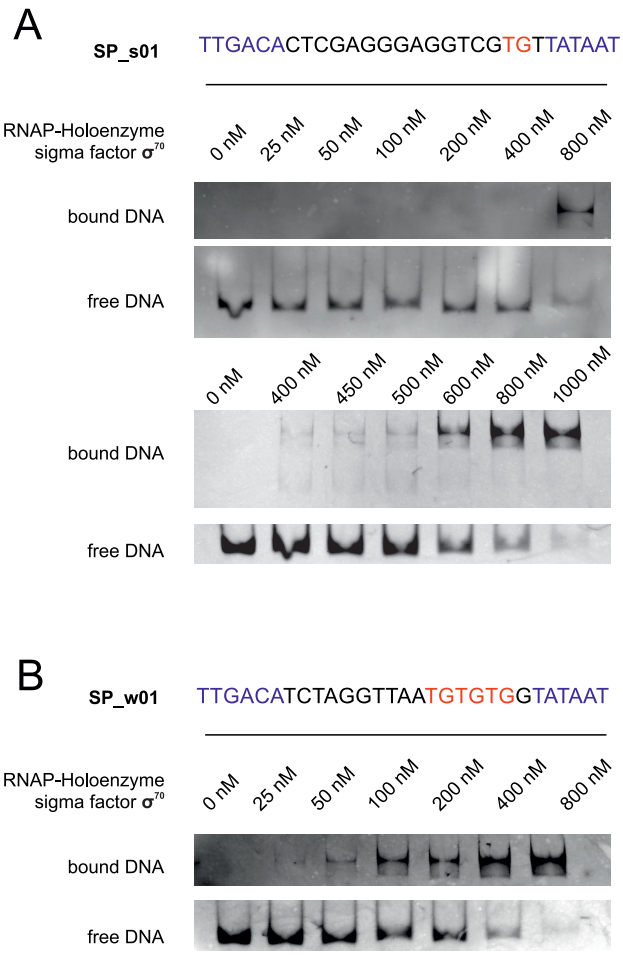


Figure 2.12: Electro Mobility Shift Assay of RNAP holoenzyme with linear DNA containing the reference promoter sequences. Electro Mobility Shift Assay was performed using cy3-labelled DNA of the strong reference promoter SP_s01 (**A**) and the weak reference promoter SP_w01 (**B**) with RNA-Polymerase enriched with σ^{70}. The DNA concentration was 4 nM, and increasing amounts of RNAP were added to the binding reaction (methods section 4.3.1). KD is calculated from the half maximal binding and is 700 nM for the strong and 100 nM for the weak promoter.

2.3 Investigating TCDS by Implementing Optogenetics to the pNovoScreen

Changing the global DNA-supercoiling of *E. coli* with sub-lethal novobiocin is straightforward and is an established method. However, novobiocin did have some unwanted effects in the NovoScreen, and we thus wanted to make our screen independent of supplementing drugs. Further, the addition of novobiocin alters the global supercoiling levels. Thus and also to investigate TCDS, a controllable neighbouring transcription on the plasmids was needed. As described by Liu and Wang [1987] the transcription alters the supercoiling in its vicinity. This local changes in supercoiling are used by previous work of, e.g. by Dages et al. [2018] and Zhi et al. [2017], to investigate the influence of neighbouring transcription on specific promoters. We wanted to explore if we could establish an inducible promoter that works without supplementation of an inducer chemical to investigate TCDS on our promoter library. Therefore, it was decided to implement an optogenetic regulated promoter that is independent of supplementation of any chemical inducer that could, in principle, interfere with the metabolism of the cell. We used the ColE promoter with LexA operator sites and the LexA-VVD repressor fusion (LEVI) as described in [Chen et al., 2016]. The promoter was cloned in front of the mCherry gene and the gene expressing LEVI on the plasmid (fig. 2.14A). This setup allowed reduction of mCherry expression up to one-tenth when cells were exposed to blue light (fig. 2.13B). This reduction of neighbouring expression should theoretically reduce the negative supercoiling produced by TCDS towards our promoter library as described in [Dages et al., 2018]. The cloning site for our promoter libraries in front of the mVenus reporter remained the same as in the pNovoScreen. In the pOptoScreen plasmid, a 17 bp library was cloned just like in the novobiocin screen (see fig. 4.2). A prototype lid with 96-LEDs the same size as the wells of a 96-well plate was used to expose cells to blue light (fig 2.13A, [Sobetzko, unpublished]). The optimal light intensities were investigated empirically with the prototype of the LED-lid (fig. 2.13 C and D, [Weile, 2018]). When 17 bp spacer library promoters were tested in our assay, they reacted to the downregulated expression of neighbouring genes by upregulating their own expression (fig. 2.14B). This seemed to be dependent on the strength of the promoter, as the upregulation was lower the stronger the promoter was (fig. 2.14B). The upregulation was dependent on the downregulation of the mCherry

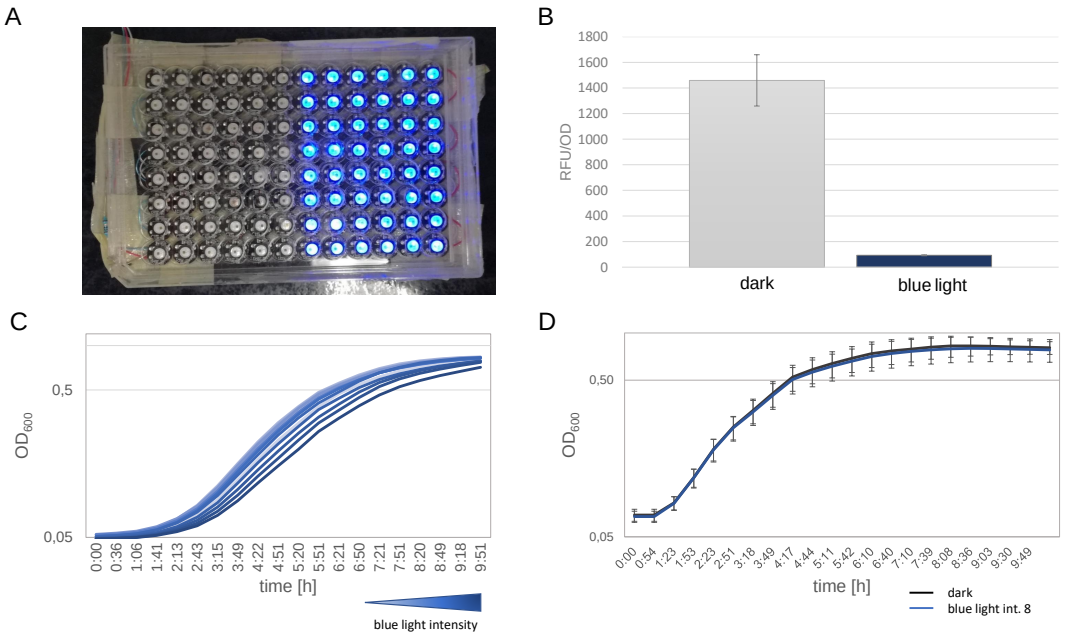


Figure 2.13: A The prototype of a LED-lid with 96-LED that are individually programmable [Sobetzko, unpublished]. **B** Dynamic range of the LEVI-LexA *ColE*-promoter controlled by blue light. **C** mean growth curves of three replicates of Top10 cells harbouring the pOptoScreen plasmid. The darker the blue line, the higher was the blue light intensity used for the test. **D** Growth of Top10 cells harbouring OptoScreen not exposed to blue light (grey line) and when exposed to blue light intensity eight (blue line) showed no difference in growth.

expression, as shown for one promoter in fig. 2.14C.

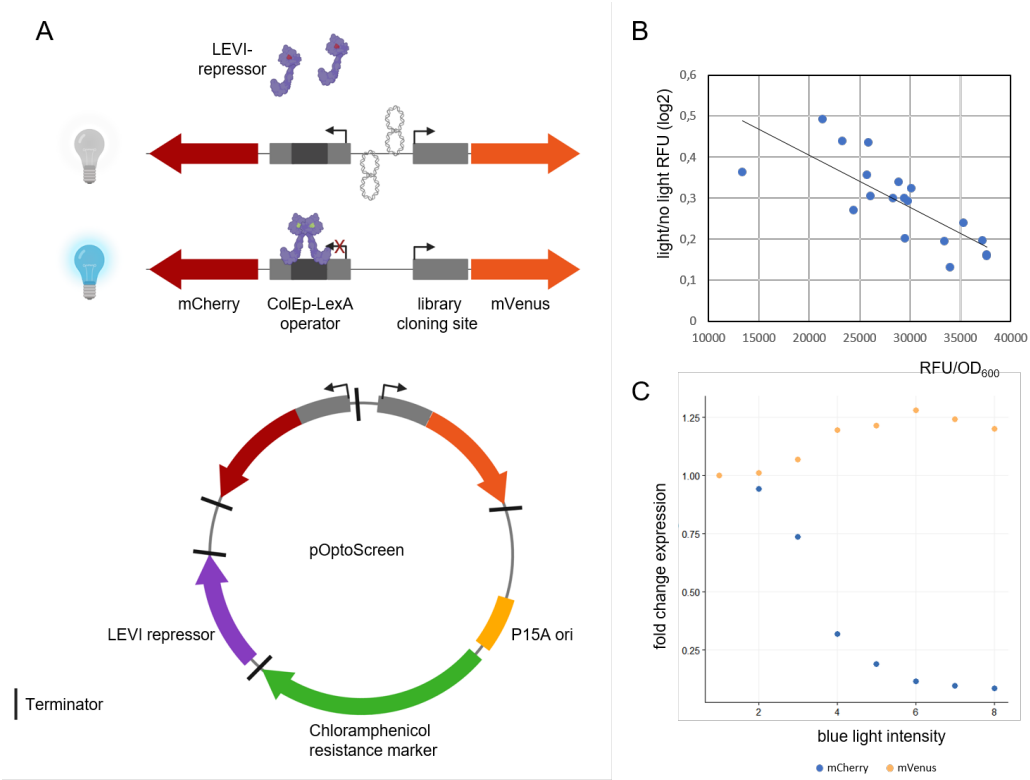


Figure 2.14: Using light switchable LexA-VVD repressor fusion (LEVI) for TCDS control. **A** The ColEp that controls mCherry expression contains LexA binding motifs in his UP-element (blue) that can be bound by the LEVI repressor. This repressor-fusion folds in when exposed to light of $\lambda = 460\,\text{nm}$ and then dimerises [Chen et al., 2016]. The cloning site for the promoter library is like in (fig. 2.6). **B** The ratio of induced vs. not induced mCherry expression against expression strength of 17 bp spacer library promoters showed a linear correlation. **C** shows a single promoter from this library. mCherry expression was lowered depending on the blue light intensity, whereas the mVenus expression was upregulated dependent on the mCherry expression.

2.4 CRISPR Swap and Drop – a Tool for Marker Free Genome Integration and Excision

For testing the findings of our TCDS screen independent from a plasmid context, it was considered whether the whole setup could be integrated into the genome. The recent method no-SCAR combined the CRISPR/Cas9 and the λ/RED [Thomason et al., 2014] system for homologous recombination [Reisch and Prather, 2015]. The idea behind this system is to cut the genome with Cas9 at the position where the insert shall recombine. Cutting with Cas9 at this position enhances cloning efficiency and makes the system independent from the cointegration of a marker gene, as when the recombination does not take place the chromosome is continuously cut by the Cas9 and the cells cannot proliferate. However, the no-SCAR system was not modular and not compatible with the standard MoClo system we use. To overcome this, we developed a set of plasmids (fig. 2.15) that makes the system modular. Further, the plasmid harbouring the part that shall integrate into the genome (pINS) is compatible with the level 1 plasmids of MoClo and the MCF-plasmids of MoCloFlex. Cloning with this system and also the combination with the modules of pSwap is possible by restriction-ligation in one-pot reaction since cloning is based on type IIs restriction enzymes BsaI and BbsI. Thus all plasmids were cleared of these type IIs restriction sites. As a constraint, all fragments that were to be used also had to be cleared of these sites.

The system also theoretically solved the problem of some parts of the chromosome are not being accessible to integration in the existing available systems. In order to be able to access essential parts of the genome, two homologous regions can be placed up and downstream of the insert DNA. By plasmid design, and by cutting the genome and the plasmid, the plasmid part can integrate into the genome, whereas the chromosomal part can integrate into the plasmid (fig. 2.16A); i.e. the parts get 'swapped'. Two additional homologies allow the swapped part on the plasmid to be integrated elsewhere, e.g. in a new strain. This feature is not so crucial to the TCDS pipeline but could be used for building large constructs in a cell, making plasmid extraction obsolete, especially if the system could be made conjugatable.

2.4.1 CRISPR Swap and Drop Plasmids and their Features

The central part of this cloning system is the pSwap. This plasmid is assembled from 7 modular plasmids in a one-pot restriction ligation (fig. 2.15). The seven modular plasmids are specific for a particular feature of the system. They all have an integration cassette just like in the MoCloFlex system harbouring a *lacZα* for blue-white screening and a *ccdB* for enhancing the cloning efficiency [Schindler et al., 2016]. Guide RNAs which will cut twice at the locus of integration are then cloned into the pT1 and pT2. pH1/Hα and pH2/Hβ harbour the homologies for swapping the chromosome part onto the pSwap (H1, H2) and to possibly drop this part at a second locus in a second strain (Hα and Hβ). Plasmids pHA and pHB contain the homologies to recombine whatever was stored in pINS into a certain locus. The backbone (origin of replication and resistance marker) of the pSwap is on pH2/Hβ.

2.4.2 Two Modes of Action: 'Swapping' and 'Dropping'

Our system has two modes of action; firstly, the swapping of a chromosomal section for a DNA fragment on the pSWAP and secondly, the integration ('drop') of the first chromosomal section into another chromosome elsewhere. For swapping, the homologies H1, HA, H2, and HB are used as depicted in figure 2.16 a. This mechanism requires the helper plasmid pCR2 which encodes the λ/RED proteins and parts of the CRISPR/Cas9 system.

For 'dropping', the plasmids pCR2, pSwap' and a plasmid called pDrop, which harbours gRNA T3 to cut the locus where the chromosomal section in pSwap' is to be inserted, are required. The homologous recombination needs the homologies Hα and Hβ (fig. 2.16 B).

To increase the efficiency of the cloning process, we implemented the violacein pathway separated on the plasmids pT1, pT2, pH1/Hα and pHA. When the pSwap is assembled, and thus all four genes *vioA*, *vioB*, *vioC*, and *vioE* are present in the pSwap, cells

harbouring this plasmid will produce deoxyviolacein and appear violet (fig. 2.17). When violet cells were observed, more than 95% of their isolated plasmids were assembled correctly, as was proven by sequencing. The four genes of the violacein pathway were derived from one operon. When separating them for our system, they had to be under the control of single promoters. When we first tried this, we observed that the violet colonies were tiny, probably because of toxicity by the stoichiometry of the four gene products. To overcome these toxic effects, we cloned a random sequence instead of their RBS and chose clones that showed no growth defects. As this system did not use a marker for integrating into the genome but was dependent on an inducible CRISPR system a visual feedback was implemented to see whether the induction, and thus the integration, had worked. Therefore, after induction and during swapping, the pSwap was engineered to also cut out the *vioC* gene (fig. 2.17). This DNA fragment harbouring the *vioC* gene is then degraded by the cell, leaving cells with a pSwap‘ which still harbours *vioA*, *vioB* and *vioE*. Together, the three encoded enzymes produce prodeoxyviolacein, a precursor to deoxyviolacein that appears greenish. Thus if after induction, green cells appeared, the homologous recombination was likely to have happened.

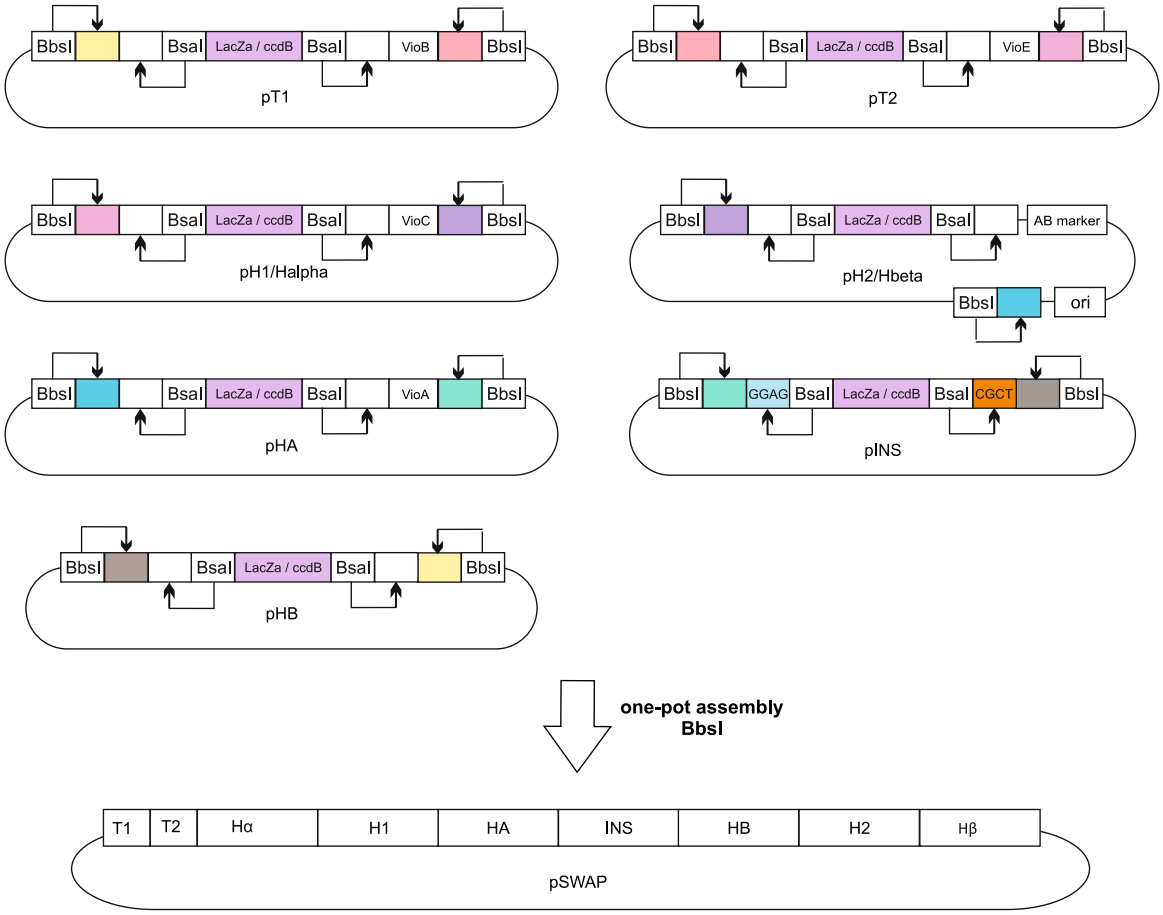


Figure 2.15: Overview over the single plasmids that can assemble to form a pSwap plasmid by a one-pot restriction ligation. To give the right order of the parts the overhangs created by BbsI are overlapping (colored boxes). All plasmids can be filled by one-pot restriction-ligation as described in fig. 2.2. pT1 and pT2 are filled with the guide RNAs that lead the Cas9 to cut in the genome, pH1/Hα harbours the left homology H1 (chromosome into pSwap) and a potential homology Hα that is important for the "dropping." Plasmid pH2/HHβ the corresponding right border. pHA and pHB harbour the homologies for the integration into the chromosome (pSwap into chromosome). The pINS harbour whatever shall be inserted into the chromosome. pINS can take every insert formed by the assembly of a MoClo level 0 library or from a MoCloFlex Destination vector.

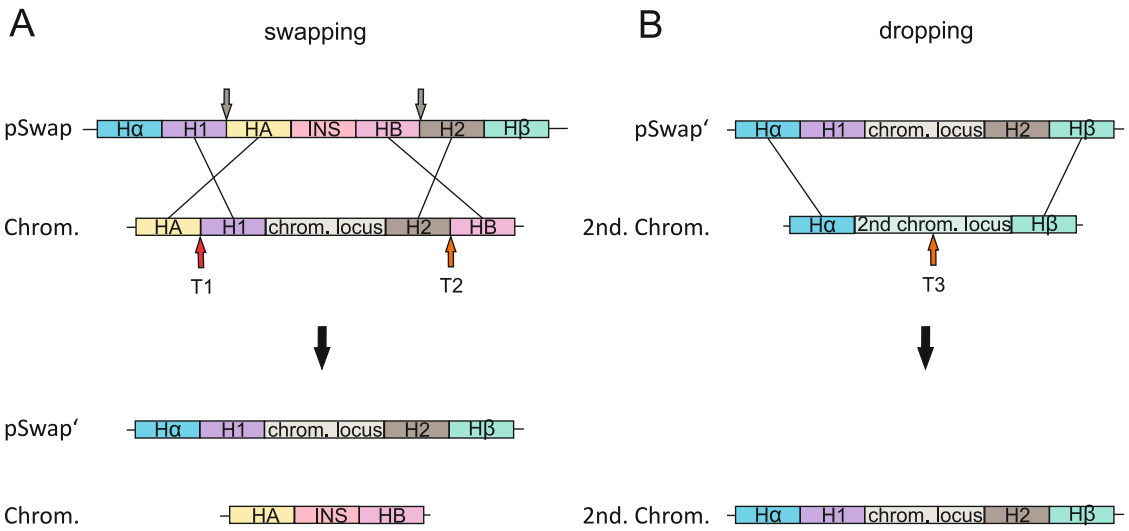


Figure 2.16: A When a pSwap is transformed into a strain harbouring the CRISPR/-Cas9 and λ/RED system on a second plasmid (pCR2), it can swap a part with the chromosome. gRNAs T1 and T2 guide Cas9 to cut the chromosome and homologous recombination occurs between between H1 and H2, while HA and HB swap the INS part with whatever part of the chromosome was chosen. **B** The resulting pSwap‘ plasmid from **A** can be transformed into a second strain, and, with the help of another plasmid (pDrop) which carries gRNA T3, drop the chromosomal section into the locus cut by T3 in the second strain. Homologous recombination for the ’dropping’ takes place between Hα and Hβ.

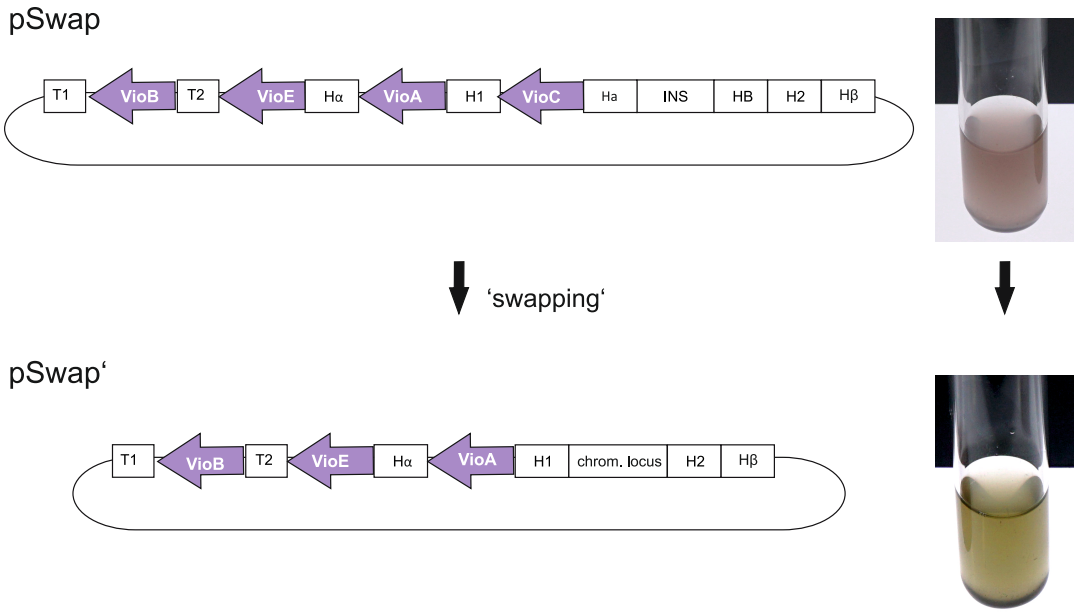


Figure 2.17: The visual assembly control feature of pSwap. to visually observe that the seven plasmids were assembled pT1, pT2 pH1/Hα and pHA carry four genes of the violacein pathway that catalyse the production of violacein that makes the colonies that have an assembled pSwap appear violet. During the swapping process the pSwap gets cut; thereby, the *vioC* is cut out. leaving the three other genes in the pSwap'. If *vioC* is missing, prodeoxyviolacein a precursor of violacein, is produced which makes the cells appear green, indicating a successful swap process.

2.4.3 Using CRISPR Swap and Drop for Creation of MG1655Δ*lac*::I-SceI

The first test of our system, mainly for checking the λ/RED and CRISPR/Cas9 induction and also the pSwap assembly was carried out in MG1655, where we attempted to enable MG1655 for counter selection through meganuclease I-SceI expression (see MoCloFlex). The meganuclease gene for I-SceI under the control of a constitutive gap-promoter was cloned into pINS. The 50 bp homology in pHA was from the *cynX* gene upstream of the *lac*-operon in pHB and was 50 bps from the *mhpR* gene downstream of the *lac*-operon. The homologies H1 and H2 were 50 bp of the left and right border of the *lac*-operon. The swapping resulted in a pSwap' harbouring the whole *lac*-operon and in *MG*1655Δ *lac*::I-SceI. The strain was cleaned from the plasmids by growth without selection pressure.

3 Discussion

3.1 The Cloning Framework of the Experimental Pipeline is Highly Compatible

The main goal of this work was to establish a pipeline to investigate position effects in transcription and the supercoiling sensitivity of promoters. For this experimental pipeline, two modular cloning systems and two promoter screening setups were developed. Accordingly, one goal was to make all these systems compatible with each other and the existing standard MoClo system. Synthetic biology consists of building functional units like artificial genetic circuits, synthetic chromosomes, or even synthetic organisms. Synthetic biology, therefore, applies engineering principles to molecular biology. The 'engineering' of biology describes the attempt to standardise and modularise biological building blocks (so-called 'bio bricks') like promoters, RBSes, or terminators, in order to use them as modules. In the past ten years, lots of building block libraries, also called part collections, were published, providing standardised parts for different model organisms. To further ease the cloning, standard cloning protocols have been suggested, where frameworks of plasmids which can be used in a modular manner exist for different part libraries.

To make use of this previous standardisation and modularisation, and to make sure our findings can be integrated easily into existing standard protocols, compatibility with the prominent modular cloning standard was implemented. In doing so, this allows the use of any part collections stored in level 0 plasmids of the MoClo protocol to build transcription units into MoCloFlex. A user can apply MoCloFlex to build networks into the MCF-Destination using different transcription units and make use of the flexibility introduced by the MCF-Linkers. It is possible to go iteratively from

MCF-Destinations back into MCF-Destination to build bigger networks or into any MoClo level 1 plasmid to use the MoClo protocol for integrating whatever build with MoCloFlex in, e.g. synthetic chromosomes. Alternatively, the content of MCF-Destinations can be built into the pINS plasmid of the CRISPR Swap and Drop system and subsequently integrated into the genome at a locus that can be freely chosen the same is possible from MoClo into the pINS (see fig. 3.1).

Compatibility, in this case, means that restriction sites used in both the systems to fill a plasmid (entry motifs) or to build into the next level (exit motifs) are shared between all parts of our experimental pipeline thus creating a framework that is also compatible to MoClo.

However, due to the differences that genomic contexts can have on transcription, and the effects that certain different plasmids may have had on the outcomes of the experiments, it was planned such that the whole setup of the pNovoScreen and pOptoScreen can be integrated into the genome. The ideal case would be first investigating the supercoiling sensitivity of a promoter at the plasmid level, before turning on the CRISPR/-Cas9 and the λ-RED systems in order to allow integration of the whole promoter library setup into the genome and running an ensuing experiment to compare the effects of promoters in a plasmid versus a genomic context.

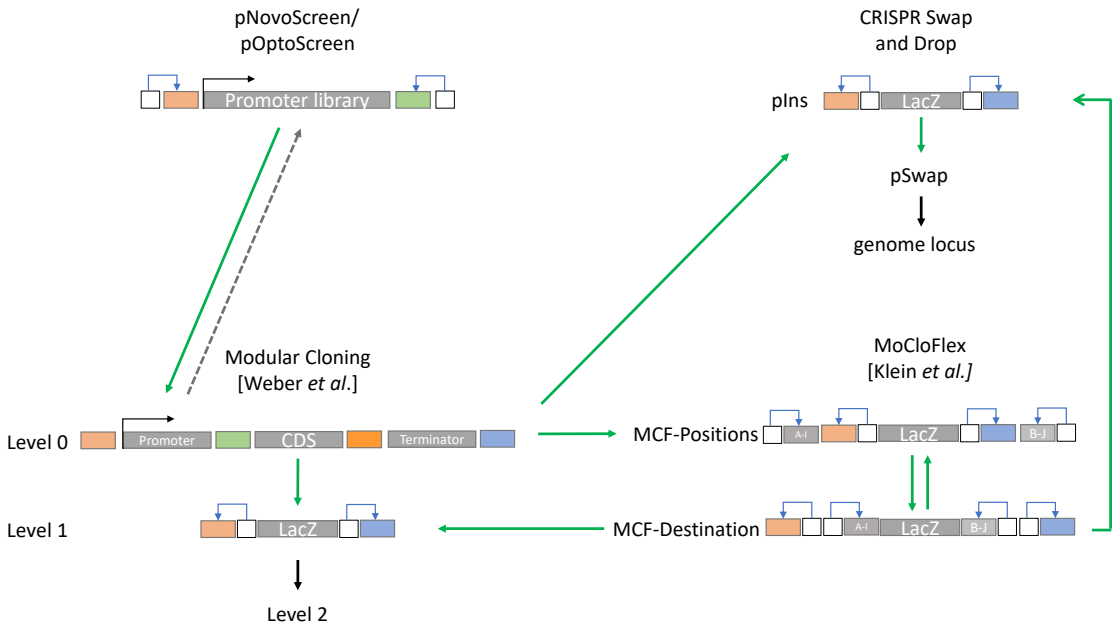


Figure 3.1: Overview of the compatibility of the different parts of this work with each other and MoClo. A green arrow indicates that the parts can be used in one-pot ligation restriction reactions. The dashed line indicates that promoters stored in MoClo vector pICH41233 can be cloned into the pOptoscreen/NovoScreen but not in a one-pot restriction ligation.

3.2 MoCloFlex Facilitates Gene Arrangement Studies

MoCloFlex is Compatible to MoClo

Up until now, a flexible cloning system which could be used to investigate the effects of gene arrangements of transcription did not exist. Our goal was to create a system which could do this and make it compatible with the MoClo protocol. Compatability was sought to make use of parts stored in level 0 plasmids in the MoClo system and to ease the exchange in the synthetic biology community. We achieved compatibility by using the same entry motifs of level 1 MoClo plasmids for the MCF-position vectors. Thus, any level 0 part library using the MoClo sites, e.g. EcoFlex [Moore et al., 2016] can also be cloned into MoCloFlex. Further, once built into the MCF-Destination, the

MCF-End-Linker can add exit motifs compatible to the MoClo level 1 entry motifs to the construct. In MoCloflex, this was able to allow iterative rounds of flexible cloning, as well as making it possible to switch back into the MoClo level 1 vector. Thus, compatibility was achieved from MoClo level 0 into MCF-Positions and from MCF-Destination into MoClo level 1. Using these motifs makes our system compatible with all systems using the MoClo protocol, e.g. systems to build synthetic chromosomes [Schindler et al., 2016; Messerschmidt et al., 2016; Zumkeller et al., 2018], systems for *E. coli* pathway optimisation [Moore et al., 2016; Vecchione and Fritz, 2019] and of course MoClo itself [Weber et al., 2011]. Therefore, MoCloFlex can use parts from these systems, but also allows these other systems the possibility to use MoCloFlex when flexibility is needed.

MoCloFlex Reduces the Cloning Effort

Transcription influences the expression of neighbouring genes. However, testing of the transcription effects of transcription units in any sequence and orientation with the existing MoClo protocol requires massive cloning effort. With MoCloFlex, we were able to achieve a significant reduction of constructs that have to be cloned, which reduces not only cloning effort but also reduces plasmid storage needed (see fig. 2.1 for comparison of MoClo and MoCloFlex cloning effort). The reusability of parts reduced space consumption for cloning significantly as all custom plasmids could be synthesised from a small set of plasmids. Moreover, as shown by Ortiz et al. [2017], MoClo (and thus MoCloFlex), workflows can be automated. Hence, even in an automated laboratory, where cloning effort may be neglectable, the reduction of storage needed makes it more efficient to use MoCloFlex for systematic gene arrangement studies. However, as MoCloFlex is dependent on type IIs restriction enzymes that have a hexameric recognition site, as are all Golden Gate-based cloning methods, parts that should enter the system often have to first be freed of those cut sites. Nevertheless, once a part is freed of these cut sites and is stored in the system, no further sequencing or manipulation of the part is needed; thus while the part library is growing the cloning expenditure is shrinking over time. One could argue that synthesis of the vectors is even more automatable and that in the near future, cloning could be outsourced to companies. However, since the prices for synthesis dropped on average to 9 ct/bp, a

construct ordered today (First quarter 2020) still needs 10–20 business days to arrive. We have shown that a custom plasmid can be planned, built, and isolated within 24 *h* using MoCloFlex.

MoCloFlex a Tool for Transcription Unit Arrangement

MoCloFlex was built to investigate the effects of transcription unit arrangements on their transcription. Accordingly, the arrangement of three transcription units was shown to make a difference in transcription. In our example transcription network (fig 2.5), the three transcription units were controlled by promoters that appeared in the Blot et al. [2006] data set as supercoiling sensitive. Thus, it is reasonable to infer that the arrangement of transcription units could be due to transcription-coupled supercoiling. Indeed different arrangements caused different expression levels of the transcription units. However, the expected TCDS for the *gypBp* promoter in the convergent orientation would be positive supercoiling. In the divergent orientation, it would be negative, and in the tandem orientation, the supercoiling from up and downstream of *gyrBp* should rule each other out. Since *gyrBp* controlled expression is lower in convergent and divergent orientations and higher in tandem, it could be possible that *gyrBp* was downregulated by both negative and positive supercoiling. For the *aldAp* controlling mCherry and *aldAp* controlling mVenus, their expression was most active in tandem orientation. Thus, no clear rule can be applied for *aldAp* since it could react both positively and negatively to either positive or negative supercoiling. Since all promoters are the strongest in the tandem orientation, it might also be that the plasmid is less supercoiled when there is less convergent or divergent transcription ongoing, which would lead to increased transcription in general. Moreover, since the plasmid has a ColE1 origin of replication that is replicated by transcription, the tandem orientation of the replicons could possibly also lead to higher copy numbers, which would in turn generally increase the expression signal since the data cannot be corrected for copy number effects. Another constraint for this plasmid-based test is the antibiotic resistance cassette present on the plasmid backbone, which also has a directional transcription and produces TCDS. All this together allows the possibility of determining the best orientation of a network of transcription units to get the highest

expression but is not straightforward when extracting rules for individual promoters. To overcome this in the future, arrangements on the MoCloFlex plasmid could also be tested in a chromosomal context. The most desirable scenario would be the insertion of the MoCloFlex arrangements into a part of the chromosome where neighbouring areas endogenously produce little to no gene expression, thus minimising external TCDS affecting the arrangement network. Another possibility could entail enlargement of the plasmid by 10–15kb of untranscribed regions, which could conceivably shield the arrangement network from supercoiling effects elicited by the plasmid backbone.

3.3 The Spacer Influences Expression Strength and Only Slightly the Supercoiling Sensitivity of Promoters

The literature search revealed the promoter spacer region as a *bona fide* target for examining TCDS sensitivity or supercoiling sensitivity of promoters. Findings of Aoyama and Takanami demonstrated that the spacer length induces supercoiling sensitivity [Aoyama and Takanami, 1988]. However, previous research that looked at the spacer focussed on one promoter at a time, usually investigating a defined spacer sequence. With our screen, we wanted to randomise spacer length and sequence to investigate the supercoiling sensitivity of resulting promoters, in order to obtain a more general knowledge of how promoter sequence and length could influence supercoiling sensitivity. One concern we had was that context in which endogenous promoters exist are so diverse that uncovering general rules of promoter-based supercoiling sensitivity would be challenging. Thus, we decided to create a minimal promoter consisting solely of the consensus -10 and -35 sequences and a fixed discriminator on a defined plasmid system. In our setup, a minimal (synthetic) σ^{70} promoter comprising only the consensus sequences plus random spacer was was sufficient for measurable promoter expression.

From our data, it would seem that differences in the spacer of a promoter alone is insufficient to explain the supercoiling sensitivity of a promoter fully. In our experi-

ment, the spacer had only a small influence on supercoiling sensitivity, far from the 6–18-fold changes that supercoiling can cause in activating the $Leu_{500}p$ shown in [Zhi et al., 2017] or in inactivating *gyrBp* [Dages et al., 2018]. However, we were able to confirm that the promoter spacer impacted expression strength, which confirms previous findings as reviewed in [Hook-Barnard and Hinton, 2007]. Nevertheless, spacer did change the supercoiling sensitivity of promoters leading to small changes in gene expression when exposed to DNA relaxation. Intriguingly, contrary to findings also showing upregulation of gene expression through DNA relaxation, [Blot et al., 2006; Ó Cróinín et al., 2006] a vast majority of promoters in our screen were down-regulated by relaxation induced with novobiocin. Further, our data revealed an expression strength-dependent response on relaxation. When the data was corrected for that effect, both up and down-regulation by relaxation could be observed; however, these effects were less intense and still revealed a tendency to more genes being downregulated by relaxation. Another hypothesis we tested was that shorter spacers should be upregulated by relaxation, whereas longer ones or optimal spacers should be downregulated since the phasing of the -10 and -35 should in the case of shorter spacers get more and in the case of longer spacers less optimal. The length of the promoter showed an influence on the expression strength but not a definite effect on the supercoiling sensitivity; thus, we could not confirm this hypothesis. Nevertheless, it could well be that the overlaying effects of expression strength in combination with novobiocin could be too strong to see other effects.

The question of what caused the use of novobiocin in our system to engender mostly negative expression thus remains. One possible explanation could be that novobiocin affects transcription elongation by blocking gyrase. During elongation, the RNAP produces supercoiling that has to be resolved by topoisomerases. Work of Chong et al. showed that the accumulation of supercoiling explained so-called transcriptional bursts. This means that if the equilibrium of topoisomerases is changed, supercoiling accumulates, and transcription is lowered; only if and when the supercoils are released will transcription continue [Chong et al., 2014]. Given that we blocked a fraction of the gyrase pool with novobiocin, the effects we observed could be due the blockage of transcription elongation, which would explain phenomenon where the stronger the expression controlled by a specific promoter from the screen, the stronger the effects of novobiocin on these promoters. However, this would also

suggest that weaker promoters would be less affected by elongation and more affected by promoter binding/initiation. Accordingly, the results showed that the weaker promoters, mainly the ones with suboptimal spacer length 12-14 bp and 21-23 bp, showed a more diverse pattern of up and downregulation even in the noncorrected data.

3.3.1 Repetitions of a 3′-TG-5′-Motif in the Spacer Positions 13–16 Increase Affinity of RNAP-Holoenzyme to Promoter

When investigating the expression strength of promoters with different spacer lengths, we observed a discrepancy between the endogenous and synthetic 17 bp spacer promoters in comparison to 18 bp promoters. It seemed that there was a cutoff on maximal expression in the the screen, raising the question of why the 17 bp spacer was not stronger than the 18bp spacer. To answer this, and also to have an idea of what makes the synthetic promoters display different levels of activity, we sequenced we sequenced 96 promoters of the highest and lowest expression quantile. Sequencing revealed the 3′-TG-5′-motif to be present in both groups to a significantly higher level than expected. Additionally, the less active group showed a repetitive 3′-TGTG-5′-motif next to the -10 region at spacer position 13–16. Other factors like bendability, melting energy or GC-content showed no significant difference between both groups, although possibly a slight trend (fig 2.9). In order to test this more fully further in the future, the number of promoters sent for sequencing could be increased. The significant occurrence of the 3′-TG-5′-motif adjacent to the -10 element (position 15–16) in these promoters far above the expected 25% in endogenous promoters having such an element, showed that we had inadvertently selected for promoters containing this motif. However, it was unclear how our screen could have selected a motif which is said to compensate for weak -35 regions. In turn, the question of what was causing this selection pressure in the screen arose.

By changing the RBS and using the promoters from the first screen, were able to confirm that there was a selection pressure in our screen that influenced maximal transcription and was not dependent on translation. We inferred from these results

that mRNA toxicity could be the cause of selection of promoter with repetitive 3'-TG-5'-motifs, which in this case may lower the expression of the promoters. However, the question of why there was no gradual toxic effect on plates, indicated by different colony size remains. An alternative suggestion could be that the 3'-TG-5'-motif was necessary for 17 bp promoter spacers; however, there were single promoters without a 3'-TG-5'-motif. Nevertheless, sequencing showed that the existence of the 3'-TGTG-5'-motif did profoundly affect the strength of 17 bp spacer promoters, lowering their expression when present. Sequential 3'-TG-5' at spacer positions 13–16 seemed to lower the expression even further. Thus, as the EMSA results suggested, 3'-TG-5'-motifs are likely responsible for the higher affinity of the RNAP holoenzyme to the promoter. The 3'-TG-5'-motif near to the -10 region is known as the extended -10 promoter motif, and, in the mix-and-match model for promoter parts, was shown to compensate for weak -35 motifs [Kumar et al., 1993; Burr et al., 2000]. The extended -10 interacts with the sigma factor; as such, it is conceivable that this interaction is even higher in the presence of two or more adjacent 3'-TG-5'-repeats. Together with a consensus -10 and a -35 sequence, which should also have optimal binding conditions to the sigma factor, the findings suggest that 3'-TG-5'-motifs create an overly tight binding, reducing the clearance of RNAP from the promoter. This could lead to the lower expression that was found in the promoters containing the 3'-TGTG-5'-motif. Work of Burr et al. [2000] suggested the 4–5 bp upstream of the -10 hexamer at *E. coli* promoters make a substantial contribution to promoter strength. We were able to show that this contribution is probably via enhancing the affinity to the RNAP-holoenzyme. This is likely since the sigma factor subunit σ_3 interacts with the promoter [Davis et al., 2017]. McCracken and Timms showed in their work that in lactobacillus strains, the 3'-TG-5'-motif stabilises transcription initiation and can even over-stabilise, and thereby hinder initiation [McCracken and Timms, 1999]. Given that we found that a high expression in our setup was presumably toxic for the cells, the over-representation of one or two sequential 3'-TG-5'-motifs might lower the transcription initiation. Additionally, genes with promoters containing optimal consensus -10 and -35 regions are shown to be expressed at lower rates than those with a few mismatches. The number of tested promoters in this work was probably insufficient to acquire enough information about the physical properties of promoters which make them more or less active, although we were able to show that there were no significant differences between weak and the strong 17 bp spacer promoters in terms of bendability, GC

content or melting energy. However, the discovery of additional trends would require more single promoters to be sequenced. Nevertheless, we found that the spacer is an essential part of the promoter to modulate the expression strength and that a high diversity of promoters could be obtained solely by changing the spacer length and sequence. Indeed, synthetic promoters with mutated spacing parts are gaining more and more attention in biotechnology and synthetic biology in order to fine-tune promoters for all kinds of genetic engineering [Blazeck and Alper, 2013; de Boer et al., 2020].

3.3.2 A Light Controlled Promoter for TCDS Research

As the novobiocin screen showed unwanted effects possibly caused by novobiocin having global negative effects in the cell, the idea was to replace the drug and induce supercoiling by neighbouring expression. Most available inducible promoters, like the *Lacp*, need the addition of a chemical inducer. The addition of the inducer in most of the experiments is an irreversible step, meaning that once the inducer is in the media, it cannot easily be removed. Another disadvantage of chemical inducers is the pipetting workload, especially, when different concentrations of inducer are tested on a 96-well plate. Therefore, we attempted using use a light-inducible promoter. Light and thus the promoter activity could be changed during experiments simulating different promoter types. A potential promoter system we considered was that created by Chen et al., where a *ColEp* with a lexA binding site is inhibited when a repression fusion called LEVI is activated by blue light. However, we questioned whether putting mCherry under the control of this promoter could influence the ensuing transcription of neighbouring genes, and specifically whether this could be via TCDS. Furthermore, we considered whether this setup could replace our novobiocin assay, as well as other previously published methods which rely on drug supplementation to investigate supercoiling.

We successfully modified a functional optogenetically-tunable promoter to be compatible with our screening system, with initial experiments showing promising results.

Our first data showed that the activation or repression of neighbouring transcription by blue light had a significant influence on the transcription controlled by the synthetic promoters. This effect of repression was dependent on the light intensity, which we controlled by a prototype LED-lid. Additionally, the synthetic promoters (17 bp random spacer) we tested, showed a response in correlation to their expression strength. The upregulation through fewer TCDS was stronger when the spacer library promoter was a less active one. However, the results also showed also that all synthetic promoters tested were upregulated by lowering neighbouring expression and thus lowering TCDS, which means that increasing TCDS through neighbouring expression would downregulate the synthetic promoters. That is in accordance with the work of Dages et al. who have shown that downregulation through TCDS should be possible. However, they could, in another example using $Leu_{500}p$, show that upregulation is also possible [Zhi et al., 2017]. Nevertheless, we only tested 40 promoters; thus, it could well be that testing more promoters could also lead to more diversity in the TCDS behaviour. Nevertheless, the fact that both transcription and global DNA supercoiling levels were not affected when using light-controlled promoters, unlike when novobiocin was used, is an important advantage of this screening protocol.

However, the observed tendency of promoters reacting positively to lowering neighbouring expression could also be a plasmid effect rather than reduced supercoiling from the neighbouring gene. It could be that the RNAP-holoenzyme is recruited to the optogenetic promoter but then switches to the synthetic promoters when repression occurs; thus the results observed could represent a competition effect of the two promoters. Competition would also explain the smaller positive effect seen with the stronger synthetic promoters, as the optogenetic promoter is only 1/10 as strong as the mean synthetic promoter. To test this hypothesis, it would be good also to test other spacer lengths with less active promoters. Another possibility would be to increase the distance between the induced and the tested promoters.

Nevertheless, like in the novobiocin test, the observed effects were not strong, thus confirming our previous findings that the spacer alone is likely not the main element which explains the supercoiling sensitivity of the promoters tested in the literature [Zhi et al., 2017; Dages et al., 2018].

3.4 CRISPR Swap and Drop – Scarless, Modular and Markerfree Genome Editing

To integrate our setup into the genome, we built a modular cloning tool around the no-SCAR method [Reisch and Prather, 2015] which is now compatible with MoClo and MoCloFlex. Furthermore, to allow access to parts of the genome that cannot be accessed by other recombineering methods. Therefore, we made it possible to swap a genome part on the plasmid while inserting the plasmid insert into the same genome locus. Last but not least, we added the violacein pathway as a visible marker to ease the assembly process and as a demonstrator of successful swapping.

We were able show that the idea of the modular CRISPR/Cas9 based genome editing plasmids worked efficiently. When pSWAP was assembled, the violacein pathway made coloured clones that were easy to pick without extensive screening. Plasmids from the coloured clones were additionally confirmed by sequencing and in all cases had the correctly assembled pSWAP. Furthermore, the phenomenon of clones becoming green in colour when the 'swapping' took place could be proven to work. We built the strain MG16655Δlac::I-SceI for the MoCloFlex toolbox with this system showing that CRISPR swap and drop is practicable for genome editing. However, integration of the *I-SceI* gene was tested and confirmed in the well-established locus of the *lac*-operon. It remains an open question if the swap process will allow also accessing parts of the chromosome where essential genes are by saving the chromosome part on the pSWAP plasmid. Similarly, the maximum size of the part that can be swapped is yet not determined. Future work could also involve implementing conjugation ability in the pSwap plasmid. If conjugable, this would allow a straightforward method to 'drop' the genomic locus that was 'swapped' from the genome into a locus on another strain. In particular, conjugation would be ideal when the 'swapped' part of the genome caused the pSwap' to be larger than 50 kb, which requires more sophisticated isolation and transformation techniques.

A feature of the modularisation of this genome editing tool is that by integrating into different loci on the genome, a library of homologies and inserts, that can be freely combined, is build up. Once a locus is confirmed for possible integration, the homologies

are stored on the plasmids pHA/pHB and can be used in every subsequent project, the same is true for the inserts and the guideRNAs.

3.5 Future Prospects

We could show that the promoter spacer contributes only little to the supercoiling sensitivity of our synthetic promoters. However, there are promoters in literature that show a significant regulation thorugh DNA supercoiling, thus other parts of the promoter are likely to influence the supercoiling sensitivity. In future, it could be prudent to test other parts of promoters for their influence on supercoiling sensitivity. Different UP-elements could influence the supercoiling sensitivity by harbouring binding motifs for proteins. The LacI inhibitor was shown to constrain negative supercoils showing supercoiling to modulate LacI binding [Fulcrand et al., 2016]. Moreover, the discriminator region is also of interest as it is melted during complex open formation and the melting energy is influenced by DNA supercoiling. During the systematic analysis of these parts, a part-library of UP-elements, spacers, discriminators, and -10 and -35 sequences could be built. Subsequently, combinations of these parts could be investigated for effects of supercoiling sensitivity. This data could then be used to create a mix-and-match model for supercoiling sensitivity and allow the rational design of promoters with certain supercoiling behaviour in a network of transcription. Such a network could subsequently be scrutinised with MoCloFlex.

The question also remains of whether the TCDS on the plasmid is sensed as supercoiling, or more specifically, as positive or negative supercoiling; therefore the pOpto-Screen with a forward mCherry expression (emitting positive supercoiling) could be used to compare promoter library behaviour. This would allow the comparison of positive TCDS effects to negative TCDS effects. For this analysis, the well studied *gyrBp* and $Leu_{500}p$ with highest activity under relaxation and hyper-negative supercoiling respectively [Zhi et al., 2017; Dages et al., 2018], could be studied. However, to generate more data, it would be worth trying to integrate the promoter screen into a robotic platform with the aim of physically creating more promoters for use in the analysis. Recent studies in yeast have analysed more than 100 million random generated promoters. Which was leading to significant observations about transcription factors regulating

eukaryotic gene expression – findings, the authors stated, that were not be possible using single promoter analysis [de Boer et al., 2020].

Biological research seems not far from having the knowledge to be able to build promoters with predicted behaviours base by base. Synthetic biology can be seen as a field of biology that tries to solve more than just biotechnological problems; I believe that the methods of synthetic biology could contribute to solving some of the fundamental general rules of biological systems. Analysing millions of randomly generated promoters may, in the future, make it possible to find also the hidden and specific design rules for gene expression, rules which are probably hard to determine when looking at a few endogenous promoters. Therefore, I hope the experimental pipeline shown in this thesis can further contribute to the systematic examination of every part of prokaryotic promoters, making both artificial biological systems more predictable and improving the fundamental understanding of gene expression across prokaryotes.

4 Material and Methods

4.1 Chemicals, Oligonucleotides and Reagents

Restriction enzymes, deoxyribose nucleoside triphospates (dNTP)s, T4 DNA Ligase (conc.), 1 *kb* DNA Ladder and Phusion DNA Polymerase were purchased from New England BioLabs (Frankfurt, Germany). Oligonucleotides were purchased from Sigma-Aldrich (Steinheim, Germany). Kits were purchased from Carl Roth GmbH + Co. KG (Karlsruhe, Germany) or VWR (Darmstadt, Germany). Chemicals were purchased from Carl Roth GmbH + Co. KG (Karlsruhe, Germany). Sequencing was performed by Eurofins GATC Biotech GmbH (Constance, Germany) and Seqlab - Sequence Laboratories Göttingen GmbH (Göttingen, Germany)

4.2 Analysis, purification and manipulation of nucleic acids

4.2.1 Annealing oligonucleotides

For annealing two oligonucleotides, *exempli gratia* (e.g.) to get a dsDNA insert for cloning, both oligonucleotides were diluted in annealing buffer to an equimolar concentration of 10 µM. The mixture was made in a PCR tube which was then placed into a PCR machine (Eppendorf, Nexus Eco). The annealing program started at 95 °C and gradually cooled down to 25 °C over 60 min in 5 min steps. Afterwards, the

concentration of dsDNA was determined using a NanoDrop™ photometer (Thermo Scientific).

4.2.2 Measuring DNA concentration

DNA concentrations were measured using a NanoDrop™ photometer (Thermo Scientific). 1 µL H_2O or elution buffer (depending on what liquid the DNA was suspended in) was used to clean the sample loader and another 1 µL was used to blank the photometer. In between blanking a Kimtech wipe was used to remove the liquid. Following blanking 1 µL of the sample was loaded, and the dsDNA setting was used to measure the DNA concentration and purity. Concentration below 5 ng/1 µL were considered noise.

4.2.3 Agarose Gel Electrophoresis

To check the length or abundance of a DNA fragment, e.g. for confirmation of clones with the right plasmid or to purify a DNA fragment Agarose Gel Electrophoresis (AGE) was performed. 0.6 g agarose was mixed with 100 mL 1 x Tris-borate-EDTA (TBE). The mixture was boiled afterwards using a microwave (NN-SD450W; Panasonic) until the agarose was dissolved completely. The mixture was then cast into a gel tray, and a suitable comb was installed at the notches (Serva, Thermo Scientific). The agarose was was then allowed to solidify at room temperature. When the gel had formed, the gel tray was installed into a gel chamber filled with 1 x TBE, and the comb was removed. Gels were run at 80 – 220 V for 60 – 15 min.

4.2.4 Restriction Digest

To identify plasmids or DNA fragments, analytic restriction digestions were performed. The DNA to be cut was searched for specific cut sites of restriction enzymes *in silico* using the software SnapGene (GSL Biotech LLC, USA). To prepare cut vector, or DNA fragments, preparatory digests were performed. The digestion mixtures were prepared

as in table 4.1. The reactions were carried out at 37 °C for 30 min and stopped at 80 °C fo 10 min unless the specifications of the enzyme indicated different temperatures or times.

Table 4.1: Analytical and preparatory restriction digest mixtures

Analytical		Preparatory	
Ingredient	Amount	Ingredient	Amount
Buffer 10x	2 µL	Buffer 10x	5 µL
DNA	0.5 µg	DNA	more than 1 µg
Enzyme	5 U	Enzyme	10 U/µg
H_2O	*ad* 20 µL	H_2O	*ad* 50 µL

4.2.5 PCR

DNA was amplified using PCR [Mullis et al., 1986]. For every PCR in which the DNA fragment was needed in further experiments, the Q5-polymerase was used (New England Biolabs). For colony PCR, a homemade *Thermus aquaticus* (*Taq*)-polymerase was used (Bichemistry facilities, Synmikro) The PCR cycler used was an Nexus Eco PCR-cycler (Eppendorf). Primers for PCR used during this work are listed in table 4.11. A standard PCR mix is shown in table 4.2 and the standard PCR program is shown in table 4.3. The TM of the Primers was calculated with SnapGene.

Table 4.2: Standard PCR mix

Ingredient	Concentration	Volume
Buffer	5x	4 µL
dNTP	20 mM/each	0.5 µL
deoxyribonucleic acid (DNA) template	5–25 ng	1 µL
Primer forward	10 µM	0.5 µL
Primer reverse	10 µM	0.5 µL
dimethyl sulfoxide (DMSO)	100 %	0–10 µL
Q5-Polymerase	2,000 units/mL	0.2 µL
		H_2O *ad* 50 µL

Table 4.3: Standard PCR program

Step	Temperature	Time
1.	98 °C	2 min
2.	98 °C	20 s
3.	Primer TM from SnapGene	20 s
4.	72 °C	1 min/1 *kb*
5.	72 °C	2 min
6.	4 °C	∞
repeat 2. – 4. 32x		

Colony PCR

To check whether colonies contain the right DNA sequence, e.g. after transformation of a new plasmid or after genome editing, colony pcr was performed. Therefore, a master-mix was prepared as given in table 4.4 depending on how many colonies had to be checked. This master-mix then was aliquoted á 20 µL into PCR tubes. Next, a sterile toothpick was used to pick a colony. Afterwards, this toothpick was scratched three times against the bottom of the PCR tube filled with master mix. PCR tubes were then mixed, carefully spun down and put into the PCR machine. The PCR-was performed using an Nexus Eco machine (Eppendorf). A standard protocol is given in table 4.5. After the PCR was done, 10 µL of the reaction were mixed with 2 µL 6x loading dye and checked via AGE.

Table 4.4: Master mix colony PCR; volumes given for one reaction

Ingredient	Concentration	Volume
Buffer	5x	4 µL
dNTP	20 mM/each	0.2 µL
Primer forward	10 µM	0.2 µL
Primer reverse	10 µM	0.2 µL
DMSO	100 %	0–4 µL
Taq-Polymerase	homemade	0.2 µL
		H_2O *ad* 20 µL

Table 4.5: Standard colony PCR program

Step	Temperature	Time
1.	98 °C	10 min
2.	98 °C	30 s
3.	Primer TM from SnapGene	30 s
4.	72 °C	1 min/1 *kb*
5.	72 °C	2 min
6.	4 °C	∞
repeat 2. – 4. 32x		

4.2.6 DNA purification from PCRs or restriction digests

Whenever DNA in form of plasmids or linear DNA had to be purified out of PCR mixes or out of restriction digests, the DNA Purification Kit from Sigma Aldrich was used. One reaction of the kit was used for 50 µL of a PCR mix or restriction digest. For eluting the DNA from the column, H_2O at 50 °C was used instead of using the elution buffer from the kit.

4.2.7 DNA purification by gel elution

Whenever a distinct DNA fragment had to be purified, e.g. from a restriction digest or a PCR reaction, a gel elution was performed. The DNA was separated by AGE, and the band was cut out using a microscope glass slide. Following this, the gel elution kit was used (see table 4.10) as per the specifications supplied by the vendor. For eluting the DNA from the column, H_2O at 50 °C was used instead of using the elution buffer from the kit.

4.2.8 Plasmid purification (miniprep)

Plasmid DNA was extracted from overnight *E. coli* culture with the Omega kit or the Rotiprep kit (see table 4.10) using the vendor specifications. For low and mid-copy plasmids (1–15 copies per cell, e. g. p15A origin of replication) double the amount

Table 4.6: Ligation-restriction one-pot reaction mixture for MoClo or MoCloFlex

Component	Stock conc.	Volume	End conc.
T4 Ligase Buffer	$10x$	2.5 μL	$1x$
Restriction enzyme (BsaI-HF, BbsI)	10 $U/\mu L$	0.5 μL	0.2 U
T4 Ligase	2000 $U/\mu L$	0.5 μL	40 $U/\mu L$
each DNA Fragment (plasmid or linear DNA)			40 $fmol$
H_2O		*ad* 25 μL	

of *E. coli* culture was used in one extraction. For both kits, H_2O at 50 °C was used instead of the elution buffer of the kit.

4.2.9 Restriction-ligation cloning (MoClo, MocloFLex)

For MoClo and MoCloFlex, the restriction-ligation process was performed in one step as described in [Weber et al., 2011; Engler et al., 2014] with the concentrations given in table 4.6. Stock concentrations of plasmids or DNA fragments were chosen as such, that pippeting volumes below 0.5 µL were omitted. The reaction was carried out for 5 h at 37 °C stopped for 10 min at 65 °C and 10 min at 80 °C, and afterwards stored at 4 °C until transformation.

4.3 Methods for protein analysis

4.3.1 Electrophoretic Mobility Shift Assay

To measure the interaction between a protein and a DNA fragment, Electrophoretic Mobility Shift Assay (EMSA) was performed. First, a 5 % polyacrylamide gel was poured (recipe given in table 4.7). Gels not required for use directly were stored wrapped in paper soaked in 0.5 % TBE and stored at 4 °C. Next, the binding reaction was prepared. Therefore, on ice, the protein was diluted in its storage buffer to a concentration from which working dilutions could be prepared. Similarly, the cy3-labelled DNA, BSA and salmon sperm DNA were diluted to working concentrations to avoid

Table 4.7: 5 % Polyacrylamide Gel for PAGE prepared in a 50 mL reaction tube

Ingredient	Volume
Acrylamide/bisacrylamide (40 %, 29:1)	1.5 mL
H_2O	9.1 mL
10x TBE	1.2 mL
10 % ammonium persulphate solution (APS)	200 µL
Tetramethylethylendiamin (TEMED)	10 µL

Table 4.8: Exemplary EMSA binding reaction

Ingredients	control	dilution 1	dilution 2	dilution 3
Salmon sperm DNA [2 mg/mL]	0.5 µL	0.5 µL	0.5 µL	0.5 µL
Bovine serum albumin (BSA)	0.5 µL	0.5 µL	0.5 µL	0.5 µL
Cy3-labelled DNA 8 nM	0.5 µL	0.5 µL	0.5 µL	0.5 µL
Protein 250 nM	-	1 µL	2 µL	4 µL
Reaction Buffer 5x	2 µL	2 µL	2 µL	2 µL
H_2O	6.5 µL	5.5 µL	4.5 µL	3.5 µL
Protein conc.	-	25 nM	50 nM	100 nM
DNA/protein ratio	-	1:6.25	1:12.5	1:25

pipetting volumes below 0.5 µL. The binding reaction was setup to a reaction volume of 10 µL as in the example given in table 4.8. The binding reaction was performed at 20 °C in the dark for 30 min. During the binding reaction, the empty Polyacrylamide Gel Electrophoresis (PAGE) was prerun in 0.5 % TBE buffer at 10 V, and the wells were cleaned afterwards by pipetting buffer up and down using a small pipette tip to access the wells. In the binding reaction, 2.5 µL SDS-free loading buffer was added. Subsequently, 10 µL of this mixture was added to each well. The optimal band resolution was achieved when, after adding the mixture to the wells, the PAGE was run at 10 V for 10 min and then shifted to 60 V for another hour.

4.4 *E. coli* methods

4.4.1 Preparation of chemically competent *E. coli*

Chemically competent cells were prepared using the $RbCl_2$ protocol adapted from [Green and Rogers, 2013]. Therefore, the *E. coli* strain of interest was thawed and 2 mL overnight culture was setup in Lysogeny Broth (LB) (if applicable supplemented with an antibiotic). The next morning 300 mL LB without an antibiotic was inoculated with 300 µL of the overnight culture, grown to 0.5 OD_{600} and then put on ice. The following steps were all carried on ice or in 4 °C precooled centrifuges. The culture was harvested into 50 mL reaction tubes and let to rest on ice for 10 min. Afterwards, the cells were spun down at 3000 rcf for 10 min, and all cells were resuspended in 10 mL Transformation Buffer 1 (TFB1). Cells were again spun down at 3000 rcf for 10 min and this time resuspended in 6 mL Transformation Buffer 2 (TFB2). The resuspended cells were aliquoted 200 µL in precooled 1.5 mL reaction tubes and stored at −80 °C for not longer than one year. Efficiency was checked with 1 ng of a reference plasmid which was confirmed in cases where more than 100 colonies appeared when streaking out 20 µL resupended pellet.

4.4.2 Transformation of chemically competent *E. coli*

Following the protocol of [Green and Rogers, 2013], chemically competent cells were thawed on ice for approximately 10 – 20 min. 90 µL of competent cells per transformation reaction is needed to get a sufficient amount of colonies. Next, DNA was added to the cells: for the retransformation of plasmids, 10 ng of DNA was sufficient; in case of ligation, the whole ligation reaction was used. Then, DNA and cells were mixed carefully and incubated for 10 min on ice. After the incubation, cells were heat-shocked at 42 °C for 45 s and immediately returned to ice for another 5 min. Afterwards, 1 mL of LB (stored at room temperature) was added and cells incubated in a lab shaker at 37 °C and 200 rpm for h1. Finally, cells were pelleted for 2 min at 1500 rcf, and the

supernatant was decanted. In the remaining supernatant, the pellet was suspended. Using glass beads, 20 µL, and the rest of the resuspendet cells were spread on separate agar plates containing antibiotics for selection.

4.5 Plate Fluorometry Assays

Fluorometric assays were performed using an Infinite® 200 PRO plate reader (Tecan). Therefore, plasmids were transformed into chemically competent *E. coli* strain Top10 [Green and Rogers, 2013] and incubated overnight on LB plates containing chloramphenicol ($17,5\ \mu g/mL$) for selection. From these plates 94 clones were selected and inoculated into 150 μL fresh LB medium. Cells were grown for 16 h at 37 $°C$ and cultures were shook at 200 rpm. Afterwards, OD_{600} and fluorescence of mCherry (ex: 587 nm, em: 610 nm) [Shu et al., 2006], mVenus (ex: 515 nm, em: 527 nm) [Kremers et al., 2006], and mTurqouise2 (ex: 434 nm, em: 474 nm) [Goedhart et al., 2012] was measured. Data analysis was carried out using R Studio [RStudio Team, 2015] and plots were generated using the ggplot2 package [Wickham, 2016].

4.5.1 Novobiocin screen

To investigate the relationship between DNA-relaxation and promoter strength, we used sub-lethal novobiocin concentrations. Novobiocin is known to intercalate with DNA gyrase and consequently relaxes the DNA *in vivo* [Schröder et al., 2014; Hoeksema et al., 1955; Maxwell, 1993]. Therefore, we prepared two 96-well plates for every overnight plate with clones to check for supercoiling-sensitivity. The first 96-well plate was filled with 100 µL LB + antibiotic, and the second plate was filled with 100 µL LB + antibiotic + novobiocin (concentration 17.5 µg/µL). From the overnight plate, 1 µL from each well was transferred to the equal position on the freshly prepared plates (A1 into A1, A2 into A2, etc.) using an Eppendorf 8-channel pipette (experimental scheme in figure 4.1). Finally, we added 50 µL mineral oil against evaporation. Both plates were incubated 15 h at 37 °C on a plate shaker at 200 rpm. Following this, OD_{600} and fluorescence were measured in a Infinite® 200 PRO plate reader (Tecan). The R script NovoScreen-v1 (see 4.12.1) was used to import the data from the reader output,

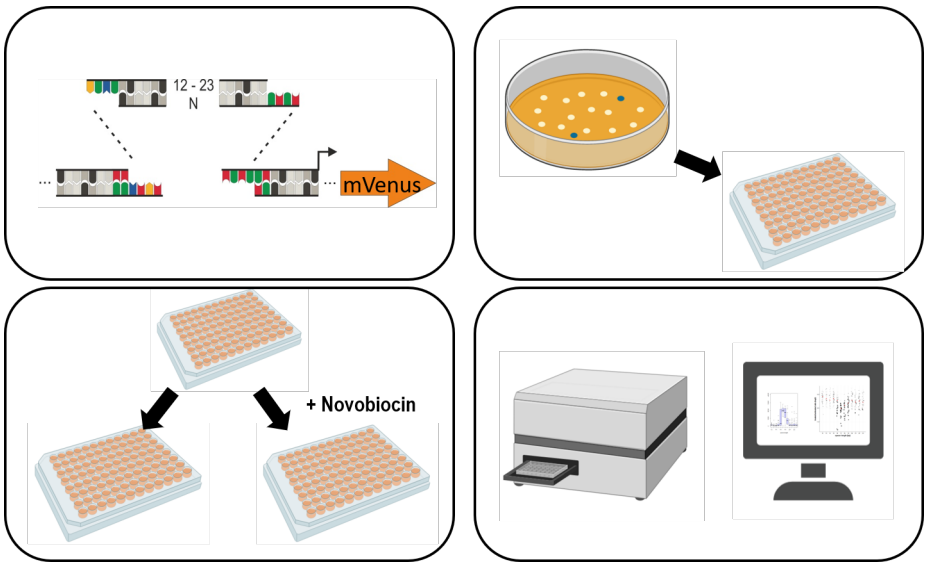


Figure 4.1: First, the library was prepared by an one-pot restriction ligation of the oligonucleotides and the pNovoScreen. The following day, blue-white screening allowed to colonies with a unique promoter to be picked and transferred into 96-well plates filled with media. The so-called master-plate was used to make an overnight culture. The following day, from this overnight plate, two new plates were inoculated following the same colony pattern. Both plates were grown overnight, one with novobiocin and one without. After After 15 h, measurements were taken and data was analyzed.

normalize fluorescence and compare the values with and without novobiocin. Clones of interest were confirmed by sequencing and further analyzed using the same R script.

4.5.2 Optogenetic screen

To investigate the influence of neighbouring expression on a promoter of interest, we built a set of plasmids which contained a light-off promoter [Chen et al., 2016] controlling the expression of the fluorescent protein mCherry next to the MoClo integration cassette. With a programmable LED lid emitting blue light, the expression could be gradually turned off (see figure 2.14). This was used to design experiments in which the expression of plasmid areas next to our promoter of interest could be gradually turned off. Since a single programmable LED shone over each well in the 96-well plate, this in theory setup the possibility to conduct 96 separate experiments

Table 4.9: *E. coli* strains used in this work

Strain	Description	Reference
DB3.1	F-, *gyrA*462, *endA*1, glnV44, Δ(*sr1-recA*), mcrB, mrr, hsdS20(rB-, mB-), ara14, galK2, lacY1, proA2, rpsL20(Smr), xyl5, Δleu, mtl1	(1993)
Top10 (Invitrogen)	F- *mcrA* Δ (mrr-hsdRMS-mcrBC) Φ 80lacZΔM15 Δ lacX74 nupG recA1 araD139 Δ (ara-leu)7697, galE15, galK16 rpsL(StrR), endA1 λ-	(1980)
MG1655	K-12 F- λ– ilvG– rfb-50 rph-1	(1980)
MG1655Δ *LacIZYA*::*I-SceI*	Δ(*lac*)::*I-SceI*	this work

on one plate. The experimental procedure was similar to the Novobiocin Screen except that on plate 2 (or one half of the plate, depending on the experiment) the LED lid was installed or the LEDs were on. First, two 96-well plates were filled with 100 µL LB + antibiotic. From the overnight plate, 1 µL from each wellwas transferred to the well at the same position of the freshly prepared plates. (A1 into A1, A2 into A2, etc.) using an Eppendorf 8-channel pipette (experimental scheme in figure 4.2). If only one plate was used, only half of the LEDs were turned on and the pipetting scheme was A1 master plate into A1 and A7, A2 into A2 and A8 *et cetera* (etc)) Following this, 50 µL mineral oil was added on top of the liquid to protect against evaporation. Both plates were incubated 15 h at 37 °C on a plate shaker at 200 rpm. After that, OD_{600} and fluorescence were measured in an Infinite® 200 PRO plate reader (Tecan). Alternatively, the plate could be incubated in the Tecan reader. Data processing was carried out using R script NovoScreen-v1 (see 4.12.1) to import the data from the reader output, to normalise fluorescence and to compare the values with and without light on.

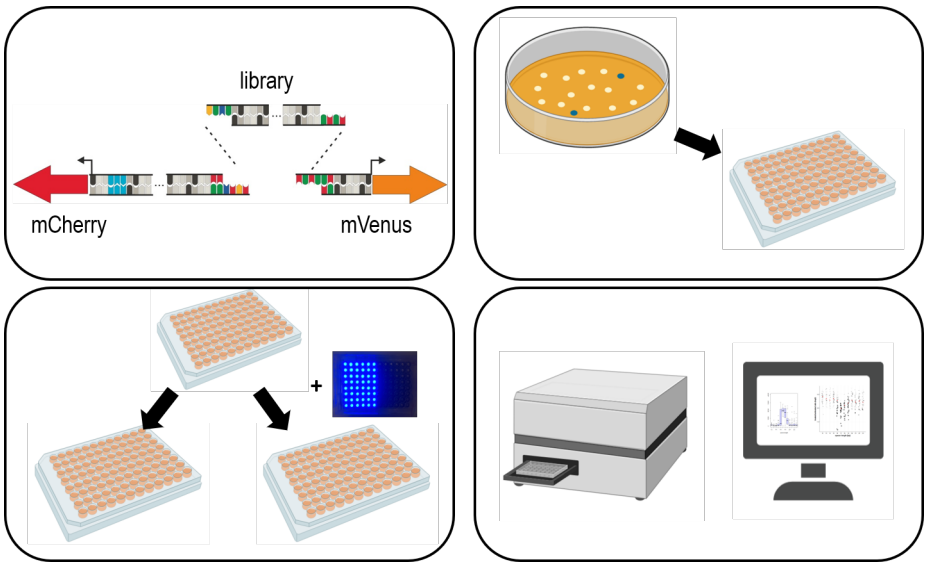


Figure 4.2: First, the library was prepared by an one-pot restriction ligation of the oligonucleotides and the pOptoScreen. The following day, blue-white screening allowed to colonies with a unique promoter to be picked and transferred into 96-well plates filled with media. The so-called master-plate was used to make a overnight culture plate. The following day, from this overnight plate, two new plates were inoculated following the same colony pattern. Both plates were grownover night, one with blue light and one in the dark. After 15 h measurements, were taken and data was analyzed.

4.6 *E. coli* strains used in this work

4.7 *E. coli* media

E. coli cells were grown in Lysogeny Broth (LB) with 1 % tryptone, 0.5 % yeast extract and 1 % sodium chloride. Solid culture medium (LA) was produced by adding 12 g/L agar-agar to LB medium. If required, LB and LA were supplemented with antibiotics at the following concentrations: ampicillin (100 $\mu g/mL$), streptomycin (100 $\mu g/mL$), kanamycin (50 $\mu g/mL$) and chloramphenicol (35 $\mu g/mL$). Liquid cultures were shaken at 200 rpm and 37 $^{\circ}C$ for all experiments.

4.8 Buffers and Solutions

Annealing Buffer	
10 mM	Tris-HCl pH 7.5
50 mM	NaCl
1 mM	Ethylenediaminetetraacetic acid (EDTA)

Loading Dye 6x	
30 % (v/v)	Glycerol
25 mM	EDTA
0.2 % (w/v)	Bromphenol blue

Lysogeny Broth - Miller formulation	
10 g/L	Bacto trypton
5 g/L	Yeast extract
10 g/L	NaCl
ad 1 L	H_2O

Storage Buffer RNAP-Holoenzyme	
20 mM	Tris-HCl pH 7.5
100 mM	NaCl
0.1 mM	EDTA
1 mM	Dithiothreitol (DTT)
50% (v/v)	Glycerol

50x Tris-acetate-EDTA (TAE) Buffer	
242.3 g/L	Tris
28.6 g/L	EDTA-Na_2 salt
60 mL	Acidic acid (100 %)
ad 1 L	H_2O
	pH 8.5

10x TBE Buffer	
107.78 g/L	Tris
7.44 g/L	EDTA-Na_2 salt
55 g/L	Borate
ad 1 L	H_2O

TFB1	
0.15 g	Potassium acetate
0.5 g	$MnCl_2$ x $4H_2O$
5 mL	1 M RbCl
0.5 mL	1 M $CaCl_2$
15 mL	Glycerol 50 % (v/v)
ad 50 mL	H_2O

	TFB2
2.5 mL	1 M Na-MOPS (pH 7)
8.75 mL	$CaCl_2$ x $4H_2O$
2.5 mL	1 M RbCl
43.75 mL	Glycerol 87 % (v/v)
ad 50 mL	H_2O
	pH 7

4.9 Kits

To prepare plasmid DNA, purify DNA from restriction digests or purify DNA after AGE, certain kits were used. Names, article number and vendors are given in table 4.10

Table 4.10: Kits used in this work

Name	Article-No.	Vendor
Roti®-Prep Plasmid Mini	HP29.2	Carl Roth GmbH + Co. KG, Karslruhe
E.Z.N.A. Plasmid DNA Mini Kit I	D6943-02	OMEGA biotek via VWR, Darmstadt
Roti®-Prep PCR purification Kit	8503.3	Carl Roth GmbH + Co. KG, Karslruhe
Roti®-Prep Gel extraction Kit	8510.3	Carl Roth GmbH + Co. KG, Karslruhe

4.10 List of oligonucleotides

Table 4.11: Oligonucleotides used in this work

ID	**Name**	**Sequences (5′ to 3′)**	**By**
CK1	fw-aldA-pICH41295	AGGGTCTCAGGAGGCTGAAACATGATGCGGTGC	C. Klein

Continued on next page

ID	Name	Sequences (5′ to 3′)	by
CK2	rev-aldA-enhanced-pICH41295	CTGGTCTCACATTTGATCTTTAGTGATAGACACCATATTGTTCCTCCTTATGGATCTGATAAATTGTTAAAGAGCGTGATTTATATGTTTGTTTTCGGTG	C. Klein
CK3	rev-aldA-pICH41295	CTGGTCTCACATTGGGCGACTCCTGTGATTTA	C. Klein
CK4	fw-mVenus-pICH41308	AGGGTCTCAAATGGTGAGCAAGGGCGAG	C. Klein
CK5	rev-mVenus-pICH41308	CTGGTCTCAAAGCTTACTTGTACAGCTCGTCCATGC	C. Klein
CK6	TopStrand-T1-Term-pICH41276	GCTTATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTAT	C. Klein
CK7	BottomStrand-T1-Term-pICH41276	AGCGATAAAACGAAAGGCCCAGTCTTTCGACTGAGCCTTTCGTTTTAT	C. Klein
CK8	fw-AraProm-pICH41295	AGGGTCTCAGGAGTTATGACAACTTGACGGCTAC	C. Klein
CK9	rev-AraProm-pICH41295	CTGGTCTCACATTTTGTTCCTCCTTATGGATCTGATAAATTGTTAAAGAGCGTGATTTATATGTTTGTTTTCGGTG	C. Klein
CK10	fw-seq-MoClo-lvl0-insert	GTTGGCCGATTCATTAATCACTC	C. Klein
CK11	rev-seq-MoClo-lvl0-insert	CTTTCGTCCACTGAAGAGCCAC	C. Klein
CK12	rev-aldA-new-pICH41295	CTGAAGACAACATTGGCGACTCCTGTGATTTATATG	C. Klein
CK13	fw-mCherry-pICH41308	AGGAAGACAAAATGGTGAGCAAGGGCGAG	C. Klein
CK14	rev-mCherry-pICH41308	CTGAAGACAAAAGCTTACTTGTACAGCTCGTCCATG	C. Klein
CK15	rev-seq-MoClo-lvl0-new	GAAAAGTGCCACCTGACGTCTAAG	C. Klein
CK16	fw-seq-MoClo-lvl0-new	CCAATACGCAAACCGCCTCT	C. Klein
CK17	fw-tufBenh-pICH41295	AGGAAGACAAGGAGCATTGTTGAGCACAATGATGTTG	C. Klein
CK18	rev1-tufBenh-pICH41295	GTTAAAGAGCTAAGACACGGATAAATCGGTGATATCACCATAGCGCGCATTCTATGGAGAC	C. Klein
CK19	rev2-tufBenh-Pich41295	AGGAAGACTACATTTGATCTTTAGTGATAGACACCATATTGTTCCTCCTTATGGATCTGATAAATTGTTAAAGAGCTAAGACACGGATAAATCGGTG	C. Klein
CK20	rev-aldA-new-pICH41295	CTGGTCTCACATTGGCGACTCCTGTGATTTATATG	C. Klein
CK21	fw-topA-HA-FRT	CTGGCTGGTCAGCATTTTATGTTGATGGCAAATGGGTTGAAGGAAAAAAATACCCATACGATGTTCCAGATTACGCTTAAGACTGGTGAGGTCTCACGC	C. Klein
CK22	rev-FRT-TopA-UTR	AGGGCCGCTTTCGCGACCCTTTGTTTATAAAAACCTGACAGAATTAAAGGCAGGGATGAGGCGCCATC	C. Klein
CK23	fw-seq-MoClo-lvl2	GGTGGCAGGATATATTGTGGTG	C. Klein
CK24	rev-seq-MoClo-lvl2	CGAACGGATAAACCTTTTCACG	C. Klein
CK25	rev-seq-start-mCherry	GGCCATGTTATCCTCCTCGC	C. Klein
CK26	fw-seq-end-mCherry	CTACACCATCGTGGAACAGTACG	C. Klein
CK27	rev-seq-start-mVenus	GACCAGGATGGGCACCACCC	C. Klein
CK28	fw-seq-end-mVenus	CACATGGTCCTGCTGGAGTTCG	C. Klein
CK29	fw-seq-MoClo-pAM60-66	CTTTATAGTCCTGTCGGGTTTC	C. Klein
CK30	rev-seq-MoClo-pAM60-66	GTTTGCAATGCACCAGGTC	C. Klein
CK31	fw-aldAmCherryT1-B3	GGAAGAAGACAATTGCAGCGATAAAACGAAAGGCCCAG	C. Klein
CK32	rev-aldAmCherryT1-B3	GGAAGAAGACAATGCCGAATTCGGATCCGGAGGCTG	C. Klein
CK33	fw-aldAmVenusT1-l3	GGAAGAAGACAATAGTAGCGATAAAACGAAAGGCCCAGTC	C. Klein
CK34	rev-aldAmVenusT1-l3	GGAAGAAGACAAGCAAGAATTCAAGCTTGGAGGCTGAAACATG	C. Klein
CK35	fw-aldAmCherryT1-A4	GGAAGAAGACAATTGCGGAGGCTGAAACATGATGCGG	C. Klein
CK36	rev-aldAmCherryT1-A4	GGAAGAAGACAATGCCGAATTCGGATCCAGCGATAAAAC	C. Klein

Continued on next page

ID	Name	Sequences (5′ to 3′)	by
CK37	fw-aldAmVenusT1-B4	GGAAGAAGACAATAGTGGAGGCTGAAACATGATGCGGTGC	C. Klein
CK38	rev-aldAmVenusT1-B4	GGAAGAAGACAAGCAAGAATTCAAGCTTAGCGATAAAACGAAAG	C. Klein
CK39	rev-seq-MoClo-pAM60-66	GACCTGGTGCATTGCAAAC	C. Klein
CK40	fw-leftendlinkertact		C. Klein
CK41	rev-leftendlinkertact		C. Klein
CK42	fw-middlelinkerccataatg		C. Klein
CK43	rev-middlelinkerccataatg		C. Klein
CK44	fw-rightendlinkeraggt		C. Klein
CK45	rev-rightendlinkeraggt		C. Klein
CK46	fw-noprom-pICH41295	GAGGTCTCAGGAGCGTAGTTGTTCGGGTGTAGTTCTGAGTAGGTGCGGTTCTAGGTCGCGTGGACTCGTGCCGCGTAGAACCGACGCTTCGGTTGCTGTGTGCGGAATGTGAGACCGA	C. Klein
CK47	rev-noprom-pICH41295	TCGGTCTCACATTCCGCACACAGCAACCGAAGCGTCGGTTCTACGCGGCACGAGTCCACGCGACCTAGAACCGCACCTACTCAGAACTACACCCGAACAACTACGCTCCTGAGACCTC	C. Klein
CK48	fw-noProm-OverlapExt	GAGGTCTCAGGAGCGTAGTTGTTCGGGTGTAGTTCTGAGTAGGTGCGGTTCTAGGTCGCGTGGACTCGTGCCGCGTA	C. Klein
CK49	rev-noProm-OverlapExt	TCGGTCTCACATTCCGCACACAGCAACCGAAGCGTCGGTTCTACGCGGCACGAGTCCACGCGACCTAGAACCGCACCTACTC	C. Klein
CK50	fw-pos1	CGCACTCTGTGAAGACAAGCCCGGAGTGAGACCCAGTGGTATGGGCAGCTGGC	C. Klein
CK51	rev-pos1	CGCACTTCGTGAAGACAAGGCTAGCGTGAGACCCACTTGAGTGGGTCACAGCTTG	C. Klein
CK52	fw-pos2	CGCACTCTGTGAAGACAAAGGCGGAGTGAGACCCAGTGGTATGGGCAGCTGGC	C. Klein
CK53	rev-pos2	CGCACTTCGTGAAGACAAGTGCAGCGTGAGACCCACTTGAGTGGGTCACAGCTTG	C. Klein
CK54	fw-pos3	CGCACTCTGTGAAGACAAGCCAGGAGTGAGACCCAGTGGTATGGGCAGCTGGC	C. Klein
CK55	rev-pos3	CGCACTTCGTGAAGACAAGTCCAGCGTGAGACCCACTTGAGTGGGTCACAGCTTG	C. Klein
CK56	fw-pos4	CGCACTCTGTGAAGACAAGAACGGAGTGAGACCCAGTGGTATGGGCAGCTGGC	C. Klein
CK57	rev-pos4	CGCACTTCGTGAAGACAATAGCAGCGTGAGACCCACTTGAGTGGGTCACAGCTTG	C. Klein
CK58	fw-MCF-linkBC	GACACTCTGTGAAGACAAAGCCGCAGCTACGCTGGTACCTGAAGGCTTGTCTTCCACGAAGTGGA	C. Klein
CK59	rev-MCF-linkBC	TCCACTTCGTGGAAGACAAGCCTTCAGGTACCAGCGTAGCTGCGGCTTTGTCTTCACAGAGTGTC	C. Klein
CK60	fw-MCF-linkDE	GACACTCTGTGAAGACAAGCACGCAGTCACGTGAGTACCACGGCCATTGTCTTCCACGAAGTGGA	C. Klein
CK61	rev-MCF-LinkDE	TCCACTTCGTGGAAGACAATGGCCGTGGTACTCACGTGACTGCGTGCTTGTCTTCACAGAGTGTC	C. Klein
CK62	fw-MCF-linkFG	GACACTCTGTGAAGACAAGGACGTCATGACGACAGTACCCGAGAACTTGTCTTCCACGAAGTGGA	C. Klein

Continued on next page

ID	Name	Sequences (5′ to 3′)	by
CK63	rev-MCF-linkFG	TCCACTTCGTGGAAGACAAGTTCTCGGGTACTGTCGTCAT GACGTCCTTGTCTTCACAGAGTGTC	C. Klein
CK64	fw-MCF-linkHA	GACACTCTGTGAAGACAAGCTAGTTCGAACTGCAGTACC GTAGCCCTTGTCTTCCACGAAGTGGA	C. Klein
CK65	rev-MCF-linkHA	TCCACTTCGTGGAAGACAAGGGCTACGGTACTGCAGTTC GAACTAGCTTGTCTTCACAGAGTGTC	C. Klein
CK66	fw-p15A-ori	CTGGTCTCTGGAGACTTCAGGTGCTACATTTGAAGAGATA AATTGC	C. Klein
CK67	rev-p15A-ori	CTGGTCTCTAGCGCGTTTTTCCATAGGCTCCGCC	C. Klein
CK68	fw-Cm-Pos2	CTGGTCTCTGGAGCAGCCTGTTGACAATTAATCATC	C. Klein
CK69	rev-Cm-Pos2	CTGGTCTCTAGCGTGATCAGCGAGCTCTAGAGAATTG	C. Klein
CK70	fw-pos1-sapI	GAGCTCTTCAAGCgaagacaaGCCCGGAGtgagacccagtggtatgggg cagctggc	C. Klein
CK71	rev-pos1-sapI	CTGCTCTTCTCACgaagacaaGGCTAGCGtgagacccacttgagtgggt cacagcttg	C. Klein
CK72	fw-pos2-sapI	GAGCTCTTCAAGCGAAGACAAAGGCGGAGTGAGACCCA GTGGTATGGGCAGCTGGC	C. Klein
CK73	rev-pos2-sapI	CTGCTCTTCTCACGAAGACAAGTGCAGCGTGAGACCCAC TTGAGTGGGTCACAGCTTG	C. Klein
CK74	fw-pos3-sapI	GAGCTCTTCAAGCGAAGACAAGCCAGGAGTGAGACCCA GTGGTATGGGCAGCTGGC	C. Klein
CK75	rev-pos3-sapI	CTGCTCTTCTCACGAAGACAAGTCCAGCGTGAGACCCAC TTGAGTGGGTCACAGCTTG	C. Klein
CK76	fw-pos4-sapI	GAGCTCTTCAAGCGAAGACAAGAACGGAGTGAGACCCA GTGGTATGGGCAGCTGGC	C. Klein
CK77	rev-pos4-sapI	CTGCTCTTCTCACGAAGACAATAGCAGCGTGAGACCCAC TTGAGTGGGTCACAGCTTG	C. Klein
CK78	rev-pos1-sapI-noccdB	CTGCTCTTCTCACgaagacaaGGCTAGCGtgagacGCGACTATG CGGCATCAGA	C. Klein
CK79	rev-pos2-sapI-noccdB	CTGCTCTTCTCACGAAGACAAGTGCAGCGTGAGACGCGA CTATGCGGCATCAGA	C. Klein
CK80	rev-pos3-sapI-noccdB	CTGCTCTTCTCACGAAGACAAGTCCAGCGTGAGACGCGA CTATGCGGCATCAGA	C. Klein
CK81	rev-pos4-sapI-noccdB	CTGCTCTTCTCACGAAGACAATAGCAGCGTGAGACGCGA CTATGCGGCATCAGA	C. Klein
CK82	fw-mTurquoise2-pICH41308	GAGAAGACAAAATGGTGAGCAAGGGCGAGGAG	C. Klein
CK83	rev-mTurquoise2-pICH41308	AGGAAGACAAAAGCTTACTTGTACAGCTCGTCC	C. Klein
CK84	fw-vio-Teil1-pICH41308	GAGAAGACAAAATGATGAGCACGTATTCTGACATTTG	C. Klein
CK85	rev-vio-Teil1-pICH41308	TCGAAGACTTATATAACTGTTCCGGAATACG	C. Klein
CK86	fw-vio-Teil2-pICH41308	CTGAAGACTAATATCGCGACTATTGGCGTC	C. Klein
CK87	rev-vio-Teil2-pICH41308	TCGAAGACTTGGTGCCGCTGTAAAACGG	C. Klein
CK88	fw-vio-Teil3-pICH41308	CTGAAGACTACACCCCGCTGCTGGGCCAG	C. Klein
CK89	rev-vio-Teil3-pICH41308	AGGAAGACAAAAGCTTAATTTACCCTTCCAAGTTTGTAC	C. Klein
CK90	fw-vioProm-pICH41295	GAGAAGACAAGGAGTGAGACGCCAGTGGTATG	C. Klein
CK91	rev-vioProm-pICH41295	AGGAAGACAACATTTGATCTTTAGTGATAGACACCATAT TGTTCCTCCTTATGGATCTGATAAATTGTTAAAGAGCTAA ACGGATACATTTACGGC	C. Klein

Continued on next page

ID	Name	Sequences (5′ to 3′)	by
CK92	fw-MCFBB-partI	GGCCTTGTCTTCCAGTGGTATGGGCAGCTGGCAC	C. Klein
CK93	rev-MCFBB-partI	GCGCAAGTCTTCCACTTGAGTGGGTCACAGCTTGTCTGTAAG	C. Klein
CK94	fw-MCFBB-partII	GAGGTCTCAGCGCTTGTTCAGAACGCTCGGTCTTGCACACCGGGCGTTTTTTC	C. Klein
CK95	rev-MCFBB-partII	AGGGTCTCAGCTGCGCCGAAAACCCCGCTTCGGCGGGGTTTTGCCGCNNNNNNNNNNTGGACTCACAAAGAAAAAACGCCCGGTGTGCAAG	C. Klein
CK96	fw-MCFBB-partIII	GAGGTCTCACAGCCTGTTGACAATTAATCATCGGCATAG	C. Klein
CK97	rev-MCFBB-partIII	CGGGTCTCATGATCAGCGAGCTCTAGAGAATTGATCCCCTCAC	C. Klein
CK98	fw-MCFBB-partIV	GAGGTCTCAATCAGGAGACTTCAGGTGCTACATTTGAAGAG	C. Klein
CK99	rev-MCFBB-partIV	CGGGTCTCATTTCCATAGGCTCCGCCCCCCTGACAAGCATC	C. Klein
CK100	fw-MCFBB-partV	GAGGTCTCAGAAAAAAAAAAAACCCCGCCCCTGACAGGGCGGGGTTTTTTTTNNNNNNNNNNTTTCTTTTGGGTATAGCGTCGTGGACAGTCATTCATC	C. Klein
CK101	rev-MCFBB-partV	CTGGTCTCAGGCCCCGGCTTATCGGTCAGTTTCACCTGATTTACGTAAAAACCCGCTTCGGCGGGTTTTTGCTTTTGGAGGGGCAGAAAGATGAATGACTGTCCAC	C. Klein
CK102	fw-MCF-linkHI	GACACTCTGTGAAGACAAGCTAGCAGCTACGCTGGTACCTGAGTACTTGTCTTCCACGAAGTGGA	C. Klein
CK103	rev-MCF-linkHI	TCCACTTCGTGGAAGACAAGTACTCAGGTACCAGCGTAGCTGCTAGCTTGTCTTCACAGAGTGTC	C. Klein
CK104	fw-MCF-linkJA	GACACTCTGTGAAGACAAGTCAGCAGCTACGCTGGTACCTGAGCCCTTGTCTTCCACGAAGTGGA	C. Klein
CK105	rev-MCF-linkJA	TCCACTTCGTGGAAGACAAGGGCTCAGGTACCAGCGTAGCTGCTGACTTGTCTTCACAGAGTGTC	C. Klein
CK106	fw-MCF-pos5IJ	GAGCTCTTCAAGCGAAGACAAGTACGGAGTGAGACCCAGTGGTATGGGCAGCTGGC	C. Klein
CK107	rev-MCF-pos5IJ	CTGCTCTTCTCACGAAGACAATGACAGCGTGAGACCCACTTGAGTGGGTCACAGCTTG	C. Klein
CK108	fw-MCFLinker-BCLong	GACACTCTGTGAAGACAAAGCCGCAGCTGACTGTAGCGACGCTCTGA	C. Klein
CK109	rev-MCFLinker-BCLong	AGCACTTCGTGGAAGACAAGCCTTCAGAGCGTCGCTACAGTCAGCTGC	C. Klein
CK110	fw-MCFLinkBC-short	GACACTCTGTGAAGACAAAGCCGCAGCTACGCTCTGA	C. Klein
CK111	rev-MCFLinkBC-short	AGCACTTCGTGGAAGACAAGCCTTCAGAGCGTAGCTGC	C. Klein
CK112	rev-LinkerA-SapI	AGGCTCTTCACACGAAGACAAGGGCTCAGAGCGTAGCTGC	C. Klein
CK113	fw-linkerB-SapI	GAGCTCTTCCAGCGAAGACAAAGCCGCAGCTACGCTCTGA	C. Klein
CK114	fw-LinkerC(rev)-SapI	GAGCTCTTCCAGCGAAGACAAGCCTGCAGCTACGCTCTGA	C. Klein
CK115	rev-linkerC-SapI	AGGCTCTTCACACGAAGACAAGCCTTCAGAGCGTAGCTGC	C. Klein
CK116	fw-LinkerD-SapI	GAGCTCTTCCAGCGAAGACAAGCACGCAGCTACGCTCTGA	C. Klein

Continued on next page

ID	Name	Sequences (5′ to 3′)	by
CK117	rev-linkerD(rev)-SapI	AGGCTCTTCACACGAAGACAAGCACTCAGAGCGTAGCTG C	C. Klein
CK118	fw-LinkerE(rev)-SapI	GAGCTCTTCCAGCGAAGACAATGGCGCAGCTACGCTCTG A	C. Klein
CK119	rev-linkerE-SapI	AGGCTCTTCACACGAAGACAATGGCTCAGAGCGTAGCTG C	C. Klein
CK120	fw-LinkerF-SapI	GAGCTCTTCCAGCGAAGACAAGGACGCAGCTACGCTCTG A	C. Klein
CK121	rev-LinkerF(rev)-SapI	AGGCTCTTCACACGAAGACAAGGACTCAGAGCGTAGCT GC	C. Klein
CK122	fw-LinkerG(rev)-SapI	GAGCTCTTCCAGCGAAGACAAGTTCGCAGCTACGCTCTG A	C. Klein
CK123	rev-LinkerG-SapI	AGGCTCTTCACACGAAGACAAGTTCTCAGAGCGTAGCTG C	C. Klein
CK124	fw-LinkerH-SapI	GAGCTCTTCCAGCGAAGACAAGCTAGCAGCTACGCTCTG A	C. Klein
CK125	rev-LinkerH(rev)-SapI	AGGCTCTTCACACGAAGACAAGCTATCAGAGCGTAGCTG C	C. Klein
CK126	fw-LinkerI(rev)-SapI	GAGCTCTTCCAGCGAAGACAAGTACGCAGCTACGCTCTG A	C. Klein
CK127	rev-LinkerI-SapI	AGGCTCTTCACACGAAGACAAGTACTCAGAGCGTAGCTG C	C. Klein
CK128	fw-linkerJ-SapI	GAGCTCTTCCAGCGAAGACAAGTCAGCAGCTACGCTCTG A	C. Klein
CK129	rev-LinkerJ(rev)-SapI	AGGCTCTTCACACGAAGACAAGTCATCAGAGCGTAGCTG C	C. Klein
CK130	fw-noProm-MCFPos	GTGGTCTCAGGAGCGTAGTTGTTCGGGTGTAG	C. Klein
CK131	rev-noProm-MCFPos	GCGGTCTCTAGCGCCGCACACAGCAAC	C. Klein
CK132	fw-MCF-EndLink-XE	GAGCTCTTCCAGCGAAGACAAGGCCGGTCTCAGGAGGC CAAAGTCTTCGTGTGAAGAGCCT	C. Klein
CK133	rev-MCF-EndLink-XE	AGGCTCTTCACACGAAGACTTTGGCCTCCTGAGACCGGC CTTGTCTTCGCTGGAAGAGCTC	C. Klein
CK134	fw-MCF-EndLink-XF	GAGCTCTTCCAGCGAAGACAAGGCCGGTCTCAGGAGGTC CAAGTCTTCGTGTGAAGAGCCT	C. Klein
CK135	rev-MCF-EndLink-XF	AGGCTCTTCACACGAAGACTTGGACCTCCTGAGACCGGC CTTGTCTTCGCTGGAAGAGCTC	C. Klein
CK136	fw-MCF-EndLink-JY	GAGCTCTTCCAGCGAAGACAAGTCACGCTTGAGACCGCG CTTGTCTTCGTGTGAAGAGCCT	C. Klein
CK137	rev-MCF-EndLink-JY	AGGCTCTTCACACGAAGACAAGCGCGGTCTCAAGCGTGA CTTGTCTTCGCTGGAAGAGCTC	C. Klein
CK138	fw-MCF-EndLink-IY	GAGCTCTTCCAGCGAAGACAAGTACCGCTTGAGACCGCG CTTGTCTTCGTGTGAAGAGCCT	C. Klein
CK139	rev-MCF-EndLink-IY	AGGCTCTTCACACGAAGACAAGCGCGGTCTCAAGCGGTA CTTGTCTTCGCTGGAAGAGCTC	C. Klein
CK140	fw-MCFBB-partI-neu	GAGGTCTCAGGCCTTGTCTTCCAGTGGTATGGGCAGCTG GCAC	C. Klein
CK141	rev-MCFBB-partI-neu	AGGGTCTCAGCGCAAGTCTTCCACTTGAGTGGGTCACAG CTTG	C. Klein
CK142	rev-p15A-ori-Seq	GATTACGCGCAGACCAAAACGATC	C. Klein

Continued on next page

ID	Name	Sequences (5′ to 3′)	by
CK143	fw-MCFbb-partORIcm	AGGGTCTCAAGTAGGTGCGGCAAAACCCCGCCGAAGCG GGGTTTTCGGCGTTCTAGGTCGACTTCAGGTGCTACATTT GAAGAG	C. Klein
CK144	rev-MCFbb-PartORIcm	TCGGTCTCACTAGAAAAAAAAAACCCCGCCCTGTCAGGG GCGGGGTTTTTTTTTGCACGAGTCCTGATCAGCGAGCTCT AGAGAATTG	C. Klein
CK145	fw-MCFbbI-partLACccdb	AGGGTCTCACTAGGTCGGCGGCAAAACCCCGCCGAAGC GGGGTTTTCGGCGCGTAGAACCGGGCCTTGTCTTCCAGTG GTATGGGCAGCTGGC	C. Klein
CK146	rev-MCFbb-partLacZccdB	TCGGTCTCATACTCAGTGGACTCACAAAGAAAAAACGCC CGGTGTGCAAGACCGAGCGTTCTGAACAAGCGGCACGA GAGCGCAAGTCTTCCACTTGAGTGGGTCACAGCTTG	C. Klein
CK147	fw-seq-cmR-end	GCTTAATGAATTACAACAGTACTGCG	C. Klein
CK148	fw-MCF-EL-X	GAGCTCTTCCAGCGAAGACAAGGCCGGTCTCAGGAG	C. Klein
CK149	rev-MCF-EL-XA	AGGCTCTTCACACGAAGACTTGGGCCTCCTGAGACCGGC C	C. Klein
CK150	rev-MCF-EL-XB	AGGCTCTTCACACGAAGACTTAGCCCTCCTGAGACCGGC C	C. Klein
CK151	rev-MCF-EL-XC	AGGCTCTTCACACGAAGACTTGCCTCTCCTGAGACCGGC C	C. Klein
CK152	rev-MCF-EL-XD	AGGCTCTTCACACGAAGACTTGCACCTCCTGAGACCGGC C	C. Klein
CK153	rev-MCF-EL-XG	AGGCTCTTCACACGAAGACTTGTTCCTCCTGAGACCGGC C	C. Klein
CK154	rev-MCF-EL-XH	AGGCTCTTCACACGAAGACTTGCTACTCCTGAGACCGGC C	C. Klein
CK155	rev-MCF-EL-XI	AGGCTCTTCACACGAAGACTTGTACCTCCTGAGACCGGC C	C. Klein
CK156	rev-MCF-EL-XJ	AGGCTCTTCACACGAAGACTTGTCACTCCTGAGACCGGC C	C. Klein
CK157	rev-MCF-El-Y	AGGCTCTTCACACGAAGACAAGCGCGGTCTCAAGCG	C. Klein
CK158	fw-MCF-EL-AY	GAGCTCTTCCAGCGAAGACAAGGGCCGCTTGAGACCGC GC	C. Klein
CK159	fw-MCF-EL-BY	GAGCTCTTCCAGCGAAGACAAAGCCCGCTTGAGACCGC GC	C. Klein
CK160	fw-MCF-EL-CY	GAGCTCTTCCAGCGAAGACAAGCCTCGCTTGAGACCGCG C	C. Klein
CK161	fw-MCF-EL-DY	GAGCTCTTCCAGCGAAGACAAGCACCGCTTGAGACCGC GC	C. Klein
CK162	fw-MCF-EL-EY	GAGCTCTTCCAGCGAAGACAATGGCCGCTTGAGACCGCG C	C. Klein
CK163	fw-MCF-EL-FY	GAGCTCTTCCAGCGAAGACAAGGACCGCTTGAGACCGC GC	C. Klein
CK164	fw-MCF-EL-GY	GAGCTCTTCCAGCGAAGACAAGTTCCGCTTGAGACCGCG C	C. Klein
CK165	fw-MCF-EL-HY	GAGCTCTTCCAGCGAAGACAAGCTACGCTTGAGACCGCG C	C. Klein
CK166	fw-gnm-intgr-SceI	CCGGTCAGGTGGTAGTTCTACTACTAGTCTCCCACATAGA TATTCCTTAGCACGTGCACAGTCGTACCTATGG	C. Klein

Continued on next page

ID	Name	Sequences (5′ to 3′)	by
CK167	rev-gnm-intgr-SceI	CCCGGGTGAGTAGTTGATACAAAACCGGCAATTTAACAA TATTTGAGCGGTGATCCACACTTGTACCAAGAGCG	C. Klein
CK168	rev-gnm-intgr-SceI	CCCGGGTGAGTAGTTGATACAAAACCGGCAATTTAACAA TATTTGAGCGGCCCGCTAAAACAATTCATCC	C. Klein
CK169	fw-BBPS-OriCm	GAGCTCTTCTCGCTCTCGTGCCGCTTGTTC	C. Klein
CK170	rev-BBPS-OriCm	CGGCTCTTCTGCTCGGTTCTACGCGCCG	C. Klein
CK171	fw-BBPS-Prom-mC	GAGCTCTTCTAGCGAAGACATGGAGTTGACACGAGACCG GTCTCGTATAATTAGCTACGATTACTAGGTCTTCGCTCTTT AACAATTTATCAGATCC	C. Klein
CK172	rev-BBPS-Prom-mC	GAGCTCTTCTGCGTTACTTGTACAGCTCGTCCATG	C. Klein
CK173	fw-PromLib14	GGTCTCGGACANNNNNNNNNNNNNTATACGAGACCG TGTCTATCAC	C. Klein
CK174	fw-PromLib17	GGTCTCGGACANNNNNNNNNNNNNNNNTATACGAG ACCGTGTCTATCAC	C. Klein
CK175	fw-PromLib20	GGTCTCGGACANNNNNNNNNNNNNNNNNNTATAC GAGACCGTGTCTATCAC	C. Klein
CK176	rev-PromLib	GTGATAGACACGGTCTCG	C. Klein
CK177	fw-ckeck-int-1,196,347	GTGACGCGATCATCGAAAAC	C. Klein
CK178	rev-ckeck-int:1,196,894	CACTGTTGGGAGTTGTAATGC	C. Klein
CK179	fw-Seq:K-12-1,196,427	GTTAGGCTGCTAACGGTTATC	C. Klein
CK180	rev-seq.K-12-1,196,788	CATTCCGCATGTACCTGAAC	C. Klein
CK181	fw-I-Sce-MC-1.2-1.3	CTGAAGACAAGCAATGCAGCGACAAATATTGATAGC	C. Klein
CK182	rev-I-Sce-MC-1.2-1.3	AGGAAGACAAGTAACTTGTACCAAGAGCGATAAAAC	C. Klein
CK183	fw-LacZ-T1	GAGCTCTTCTAGCGAAGACATGGAGTTGACACGAGACCA GTTACGCTAGGGATAACAGGGTAATATAGCAGTGGTATG GGCAGCTGG	C. Klein
CK184	rev-LacZ-T1	CGGCTCTTCTTAGCAACCGAAGCATAAAACGAAAGGCCC AGTCTTTCGACTGAGCCTTTCGTTTTATTAGAACCGCACT ATGCGGCATCAGAGCAG	C. Klein
CK185	fw-flexProm-ccdB	GAGCTCTTCTCTATAGAACCGACTTGACANNNNNNNNN NNNNNNNNTATAATGCTCTTTAACAATTTATCAGATCCA TAAGGAGTATGTCAATGCAGTTTAAGGTTTACACC	C. Klein
CK186	rev-flexProm-ccdB	CGGCTCTTCTGTCGAAGACCTAGTAATCGTAGCTAATTAT ACGAGACCCCAAACTGTCTCACGACGTTTTGAACCCAGT TATATTCCCCAGAACATCAGG	C. Klein
CK187	fw-mCherry-spacerBB	GAGCTCTTCTGACGCTCTTTAACAATTTATCAGATCCATA AGGAGGAACAATATGGTGTCTATCACTAAAGATCAAATG GTGAGCAAGGGCGAG	C. Klein
CK188	fw-H1Ha-Insert-part1	GAGGTCTCAGGAGGCTCGATCGCTAATCGTTATCGGCCC ATGCTCACCGTTAAGCAG	C. Klein
CK189	rev-H1Ha-Insert-part1	GAGGTCTCATTCCTTGTAAATCGCACTAAAG	C. Klein
CK190	fw-H1Ha-Insert-part2	GAGGTCTCAGGAACATGATGCGACGCTTGTTC	C. Klein
CK191	rev-H1Ha-Insert-part2	GAGGTCTCAACGGGTCATGTGGTTATTAATCCAGG	C. Klein
CK192	fw-H2Hb-Insert-part1	GAGGTCTCAGGAGCGTTTTCGCAGAAACGTGG	C. Klein
CK193	rev-H2Hb-Insert-part1	TCGGTCTCACATGCATTTACGTTGACACC	C. Klein
CK194	fw-H2Hb-Insert-part2	GAGGTCTCACATGCTTAAAATATCCTGATCTACCATCGC	C. Klein
CK195	rev-H2Hb-Insert-part2	GAGGTCTCAAGCGGCTCGATCGCTAATCGTTATCGGCCC GCTATTTTACCCATTG	C. Klein

Continued on next page

ID	Name	Sequences (5′ to 3′)	by
CK196	fw-Int-Part-1	GCGCTCTTCAGCGGATATGCCCGTAGCTACTAGCGGTTGTTCAGAACGCTCGGT	C. Klein
CK197	rev-Int-Part-1-RNA-swap	ATAACGATTAGCGATCGAGCGTGCTCAGTATCTCTATCACTGATAG	C. Klein
CK198	for-Int-Part-1-RNA-swap	GCTCGATCGCTAATCGTTATGTTTTAGAGCTAGAAATAGCAAG	C. Klein
CK199	rev-Int-Part-1	GCGCTCTTCAGGAACTTCAGGTGCTACATTTGAAG	C. Klein
CK200	fw-Int-Part-2	GCGCTCTTCATCCCTATCAGTGATAGAGATTGAC	C. Klein
CK201	rev-Int-Part-2	GCGCTCTTCACGGGATGTCAAGAGAGGCTTGA	C. Klein
CK202	fw-Int-Part3	GCGCTCTTCACCGCTAGTAGCTACGGGCATATCTTACCAATGCTTAATCAGTGAGG	C. Klein
CK203	rev-Int-Part-3	GCGCTCTTCACGCGGAACCCCTATTTGTTTATTTTTC	C. Klein
CK204	fw-H2Sc-Part-1	GCGCTCTTCATCAGTTTAGCTTACGCTGCAC	C. Klein
CK205	rev-H2Sc-Part-1	CTGCTCTTCTTTCTTCAAAAAACGAAGGCTCCC	C. Klein
CK206	for-H2Sc-Part-2	CTGCTCTTCTGAACTATTGACGCGAGATTATTTCTTC	C. Klein
CK207	rev-H2Sc-Part-2	GCGCTCTTCATGGTGGACTCACAAAGAAAAAACGCCCGGTGTGCAAGACCGAGCGTTCTGAACAAGATCCCGACTTTAGGCTCCAGCGGCCATC	C. Klein
CK208	for-H2CrSceI-Part-3	GCGCTCTTCACCACTATATTACCCTGTTATCCCTAGCGTAACTTCCCTCGTCTTCGAAGACATTGGTAATTGTCAACAGGCTGTTGTTCAGAACGCTCGGTC	C. Klein
CK209	rev-H2Sc-Part3-RNAint	CTAGTAGCTACGGGCATATCGTGCTCAGTATCTCTATCACTGATAG	C. Klein
CK210	fw-H2Sc-Part3-RNAint	GATATGCCCGTAGCTACTAGGTTTTAGAGCTAGAAATAGCAAG	C. Klein
CK211	rev-H2Sc-Part-3	GCGCTCTTCATCAGTGAGCGAGGAAGCGGAATATATCC	C. Klein
CK212	for-H2Sc-Part-4	GCGCTCTTCATGATCCACACTTGTACCAAGAG	C. Klein
CK213	rev-H2Sc-Part-4	GCGCTCTTCAAGGGTGAGCGAGGAAGCGGAATATATCC	C. Klein
CK214	fw-H2Sc-Part-5	GCGCTCTTCACCTTAAGGTATACTTTCCGCTG	C. Klein
CK215	rev-H2Sc-Part-5	GCGCTCTTCATGAAGCTTATCGGCCAGCCTCG	C. Klein
CK216	for-H2ScCreI-Part-3	GCGCTCTTCACCAAACTGTCTCACGACGTTTTGAACCCAGTCCCTCGTCTTCGAAGACATTGGTAATTGTCAACAGGCTGTTGTTCAGAACGCTCGGTC	C. Klein
CK217	for-targetDNAintegration	AGGGTCTCTGCACTAGCCGCTTCTCCCAAACCAGTTTAGAGACCAG	C. Klein
CK218	rev-targetDNAintegration	CTGGTCTCTAAACTGGTTTGGGAGAAGCGGCTAGTGCAGAGACCCT	C. Klein
CK219	rev-noProm-OverlapExt-lvl1	TCGGTCTCAAGCGCCGCACACAGCAACCGAAGCGTCGGTTCTACGCGGCACGAGTCC	C. Klein
CK220	fw-MCF-pos2-cmR-SceI		C. Klein
CK221	rev-MCF-pos2-cmR-SceI		C. Klein
CK222	fw-MCF-pos-LEVI	GCGGTCTCAGGAGCCTATGCAGCGACAAATATTG	C. Klein
CK223	rev-MCF-pos-LEVI	GCGGTCTCAAGCGTGGACTCACAAAGAAAAAACGCCCGGTGTGCAAGACCGAGCGTTCTGAACAAATCGCTACTATTCCGTTTCGCACTGG	C. Klein
CK224	fw-ColEp-LEX-pICH41233	GCGAAGACAAGGAGTGTTTTTTTGATCGTTTTC	C. Klein
CK225	rev-ColEp-LEX-pICH41233	GCGAAGACAAAGTACCGTACATATAAACG	C. Klein
CK226	fw-seq-lac-int-cps-53	GCAGAACGTGCGGAAAACATTAAG	C. Klein
CK227	rev-seq-lac-int-cps-53	CACTCGCATCCATCAGAAGTTG	C. Klein

Continued on next page

ID	Name	Sequences (5′ to 3′)	by
CK228	fw-optoscreen-LmCfw	GAGCTCTTCTAGCTGGACTCACAAAGAAAAAACGCCCG GTGTGCAAGACCGAGCGTTCTGAACAAAAGTCAGCCCCA TACGATATAAGAATTCC	C. Klein
CK229	rev-optoscreen-lmCrev	CGGCTCTTCTGCCAGAGCGTTCACCGACAAACAACAGAT AAAACGAAAGGCCCAGTCTTTCGACTGAGCCTTTCGTTTT ATTTGATGCCTGGCTGTGGGTAACTTTG	C. Klein
CK230	fw-optoscreen-lacZ	GAGCTCTTCTGGCGAAGACATGGAGTTGACACGAGACCA GTTACGCTAGGGATAACAGGGTAATATAGCAGTGGTATG GGCAGCTGG	C. Klein
CK231	rev-optoscreen-lacZ	CGGCTCTTCTGTCGAAGACCTAGTAATCGTAGCTAATTAT ACGAGACCCTATGCGGCATCAGAGCAG	C. Klein
CK232	fw-optoscreen-mVenus	GAGCTCTTCTGACGCTCTTTAACAATTTATCAGATCCATA AGGAGGAACAATATGGTGTCTATCACTAAAGATCAAATG GTGAGCAAGGGCGAGG	C. Klein
CK233	rev-optoscreen-mVenus	CGGCTCTTCTGCGCAACCGAAGCATAAAACGAAAGGCC CAGTCTTTCGACTGAGCCTTTCGTTTTATAAGCTTACTTGT ACAGCTCGTC	C. Klein
CK234	fw-optoscreen-LmCrev	GAGCTCTTCTAGCTTTGTTTATTTTTCTAAATACATTCAAA TATGTATCCGCTCATGA	C. Klein
CK235	rev-optoscreen-LmCrev	CGGCTCTTCTGCCGCTAGCTCACTCGGTCGCTACCTGTTTT TTTGATCGTTTTCACAAAAATGGAAGTACACAG	C. Klein
CK236	fw-backboneBeide-DraIII	AACACAGCGTGGAAGATCATCTTATTAATCAG	C. Klein
CK237	rev-BackboneAK-DraIII	CACACCGCGTGCTAAAACAATTCATCCAGTAAAAT	C. Klein
CK238	rev-BackboneCS-DraIII	CACACCGCGTGTTATTTGCCGACTACCTTGGTG	C. Klein
CK239	fw-p15A-DraIII	AGCACGCGGTGTGGATATATTCCGCTTCC	C. Klein
CK240	rev-p15A-DraIII	TCCACGCTGTGTTGAGATCGTTTTGGTCTGCGC	C. Klein
CK241	fw-OptoS-Term	CGGCTCTTCTGCATGTTTTTTTGATCGTTTTCAC	C. Klein
CK242	rev-OptoS-Term	CAGCTCTTCATGCGCGAAAAAACCCCGCCGAAGCGGGGT TTTTTGCGGGTAGCGACCGAGTGAGCTAG	C. Klein
CK243	fw-OptoS-Term-Library	CGGCTCTTCTGCATGTTTTTTTGATCGTTTTCACAAAAATG GAAGTACACAGTTGACANNNNNNNNNNNNNNNNNTA TAATTTATGCCGTATATAAAAACGGCG	C. Klein
CK244	fw-ak-cs-BB-phd	TACACTTAGTGAAGTTCTTCCGGTGATATATGCCTCATTA TC	C. Klein
CK245	rev-ak-cs-BB-phd	GCCACCTAGTGACGGCTCTCTCTTTTATAGGTGTAAAC	C. Klein
CK246	fw-ak-cs-delta-ccdB	TACACTAGGTGTATCGTCTGTTTGTGGATGTACAG	C. Klein
CK247	rev-ak-cs-delta-doc	GCCACTAAGTGATTGCGCTTCATGATGCGAATATAAGC	C. Klein
CK248	fw-backboneBeide-Esp3I	TACGTCTCAGCACAGATCATCTTATTAATCAGATAAAAT ATTTC	C. Klein
CK249	rev-BackboneAK-Esp3I	GTCGTCTCAACGTCTAAAACAATTCATCCAGTAAAATAT AATATTTTATTTTC	C. Klein
CK250	rev-BackboneCS-Esp3I	GCCGTCTCAACGTTTATTTGCCGACTACCTTGGTG	C. Klein
CK251	fw-p15A-Esp3I	TACGTCTCAACGTTGGATATATTCCGCTTCCTCG	C. Klein
CK252	rev-p15A-Esp3I	GTCGTCTCAGTGCTTGAGATCGTTTTGGTCTGC	C. Klein
CK253	fw-20spcr-9dscrm	ACGAAGACATGGAGTTGACANNNNNNNNNNNNNNNN NNNNTATAATNNNNNNNNNTACTAGGTCTTCATGCGCG ATATATCGATCGTG	C. Klein

Continued on next page

ID	Name	Sequences (5′ to 3′)	by
CK254	fw-17spcr-9dscrm	ACGAAGACATGGAGTTGACANNNNNNNNNNNNNNNNNTATAATNNNNNNNNNTACTAGGTCTTCATGCGCGATATATCGATCGTG	C. Klein
CK255	fw-14spcr-9dscrm	ACGAAGACATGGAGTTGACANNNNNNNNNNNNNNNNTATAATNNNNNNNNNTACTAGGTCTTCATGCGCGATATATCGATCGTG	C. Klein
CK256	rev-spcr-dscrmntr	CACGATCGATATATCGCGCAT	C. Klein
CK257	rev-P15A-Aknew	GTCGTCTCAGTGCGCCAAACTGTCTCACGACGTTTTGAACCCAGAAGACCACATCTGAGACCTTGAGATCGTTTTGGTCTGC	C. Klein
CK258	fw-Aknew-ampProm	GTCGTCTCAGCACGGAGTTGACAGAGTTTATATTGGTCATATAATGTACTCGGTCTTTAACAATTTATCAGATCCACGAGGAAACGTCCAATGAGTATTCAACATTTCCGTG	C. Klein
CK259	rev-Aknew	GTCGTCTCACTCCCAAATAGGGGTTCCGCGCCGGC	C. Klein
CK260	fw-Aknew-kanProm	GTCGTCTCAGGAGTTGACAGAGTTTATATTGGTCATATAATGTACTCGGTCTTTAACAATTTATCAGATCCACGAGGAAAGTCCAATGGCTAAAATGAGAATATCACCG	C. Klein
CK261	rev-p15A-Csnew	GTCGTCTCAGTGCCTATATTACCCTGTTATCCCTAGCGTAACTTCCGAAGACCACATCTGAGACCTTGAGATCGTTTTGGTCTGC	C. Klein
CK262	fw-Csnew-CmProm	GTCGTCTCAGCACGGAGTTGACAGAGTTTATATTGGTCATATAATGTACTCGGTCTTTAACAATTTATCAGATCCACGAGGAAACGTCCAATGGAGAAAAAAATCACTGGATATAC	C. Klein
CK263	rev-Csnew-Cmrev	GTCGTCTCATCACGCCCCGCCCTGCCACTC	C. Klein
CK264	fw-Csnew	GTCGTCTCAGTGAAGTTACGCTAGGGATAACAGGGTAATATAGCTGACGTGCACAGTCGTACC	C. Klein
CK265	rev-Csnew	GTCGTCTCACTCCCAAATAGGGGTTCCGCGCCGGC	C. Klein
CK266	fw-Csnew-StpProm	GTCGTCTCAGGAGTTGACAGAGTTTATATTGGTCATATAATGTACTCGGTCTTTAACAATTTATCAGATCCACGAGGAAAGTCCAATGCGCTCACGCAACTGG	C. Klein
CK267	fw-PromLib15	GGTCTCGGACANNNNNNNNNNNNNNNTATACGAGACCGTGTCTATCAC	C. Klein
CK268	fw-PromLib16	GGTCTCGGACANNNNNNNNNNNNNNNNTATACGAGACCGTGTCTATCAC	C. Klein
CK269	fw-PromLib18	GGTCTCGGACANNNNNNNNNNNNNNNNNNTATACGAGACCGTGTCTATCAC	C. Klein
CK270	fw-PromLib19	GGTCTCGGACANNNNNNNNNNNNNNNNNNNTATACGAGACCGTGTCTATCAC	C. Klein
CK271	rev-SBB-no-ccdB	GGCTCTTCTGTCGAAGACCTAGTAATCGTAGCTAATTATACGAGACCATAAAACGAAAGGCCCAGTCTTTCGACTGAGCCTTTCGTTTTATTAGAACCGCACTATGCGGCATCAGAGCAG	C. Klein
CK272	fw-PromLib11	GGTCTCGGACANNNNNNNNNNTATACGAGACCGTGTCTATCAC	C. Klein
CK273	fw-PromLib12	GGTCTCGGACANNNNNNNNNNNTATACGAGACCGTGTCTATCAC	C. Klein
CK274	fw-PromLib13	GGTCTCGGACANNNNNNNNNNNNTATACGAGACCGTGTCTATCAC	C. Klein

Continued on next page

ID	Name	Sequences (5′ to 3′)	by
CK275	fw-PromLib21	GGTCTCGGACANNNNNNNNNNNNNNNNNNNNTATACGAGACCGTGTCTATCAC	C. Klein
CK276	fw-PromLib22	GGTCTCGGACANNNNNNNNNNNNNNNNNNNNTATACGAGACCGTGTCTATCAC	C. Klein
CK277	fw-PromLib23	GGTCTCGGACANNNNNNNNNNNNNNNNNNNNTATACGAGACCGTGTCTATCAC	C. Klein
CK278	fw-SVA6-prom	GCGAAGACATGGAGTTGACAACTATTTTTGCATGTGATATAATTAGCTACGATTACTAGGTCTTCAT	C. Klein
CK279	rev-SVA6-prom	ATGAAGACCTAGTAATCGTAGCTAATTATATCACATGCAAAAATAGTTGTCAACTCCATGTCTTCGC	C. Klein
CK280	fw-PromLib-no35	ACGAAGACATGGAGNNNNNNNNNNNNNNNNTATAATTAGCTACGATTACTAGGTCTTCATGCGCGATATATCGATCGTG	C. Klein
CK281	fw-PromLib-no10	ACGAAGACATGGAGTTGACANNNNNNNNNNNNNNNNTAGCTACGATTACTAGGTCTTCATGCGCGATATATCGATCGTG	C. Klein
CK282	rev-Esp3I-cm-removal	CTCGTCTCTCACGAAAAACATATTCTCAATAAACCCTTTAGG	C. Klein
CK283	fw-Esp3I-cm-removal	GACGTCTCACGTGTCAGCCAATCCCTGGGTGAG	C. Klein
CK284	fw-PromLib-NoSigma70	ACGAAGACATGGAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTACTAGGTCTTCATGCGCGATATATCGATCGTG	C. Klein
CK285	fw-gyrA-fenfei2018	GAAGACATGGAGGGATGTGAATAAAGCGTATAGGTTTACCTCAAACTGCGCGGCTGTGTTATAATTTTAGCTACGATTACTAGGTCTTC	C. Klein
CK286	rev-gyrA-fenfei2018	GAAGACCTAGTAATCGTAGCTAAAATTATAACACAGCCGCGCAGTTTGAGGTAAACCTATACGCTTTATTCACATCCCTCCATGTCTTC	C. Klein
CK287	fw-insert-OS-fw-rev	CATCGTCTCACTTCAGTTACGCTAGGGATAACAGGG	C. Klein
CK288	rev-insert-OS-fw-rev	CGCGTCTCACTTCCTATGCGGCATCAGAGCAG	C. Klein
CK289	for-bb-OS-rev-fw	CATCGTCTCAGAAGACCTTACTGACGCTCTTTAACAATTTATCAGATCCATAAG	C. Klein
CK290	rev-bb-OS-fw	ATCGTCTCAGAAGACATCTCCGCCAGAGCGTTCAC	C. Klein
CK291	rev-bb-OS-rev	GCCGTCTCAGAAGACATCTCCGCCGCTAGCTCACTCGGTCGCTACC	C. Klein
CK292	fw-tetA	GCCGTCTCGAGCCGCGAAAAAACCCCGCCGAAGCGGGGTTTTTTGCGCTAAGCACTTGTCTCCTGTTTAC	C. Klein
CK293	rev-tetA	CCCGTCTCGGCCTATGAATAGTTCGACAAAGATCG	C. Klein
CK294	fw-tetR	CCCGTCTCGAGGCAGATCCGAAGTCCTCTTTAG	C. Klein
CK295	rev-tetR	CTCGTCTCGGGGCAAAAAAAACCCCGCCCTGTCAGGGGCGGGGTTTTTTTTTAACAAGTCTTAAGACCCACTTTC	C. Klein
CK296	fw-backbone	CCCGTCTCGGCCCCGTGCCGCTTGTTCAGAAC	C. Klein
CK297	rev-backbone	TCCGTCTCGGGCTGCGCAACCGAAGCATAAAAC	C. Klein
CK298	fw-Leu-500	GCGAAGACATGGAGCAGAAAATAGAGTTGACATTAAACGGCATATCCAGTGCCACTAATAGCTACGATTACTAGGTCTTCGG	C. Klein
CK299	rev-Leu-500	CCGAAGACCTAGTAATCGTAGCTATTAGTGGCACTGGATATGCCGTTTAATGTCAACTCTATTTTCTGCTCCATGTCTTCGC	C. Klein

Continued on next page

ID	Name	Sequences (5′ to 3′)	by
CK300	fw-OS-in-pINS	GCGGTCTCAGGAGGTTTTCGGCGCGTAGAACCG	C. Klein
CK301	rev-OS-in-pINS	GCGGTCTCAAGCGGTTCTGAACAAGCGGCACGAGA	C. Klein
CK302	Fw-OligoTestSeq	GATTCTGTGGATAACCGTATTAC	C. Klein
CK303	fw-OligoTest	GCCGTCTCATATAGGGCCTCGTGATACGCC	C. Klein
CK304	rev-OligoTest	GCCGTCTCATGTCTCGGTCGTTCGGCTGCG	C. Klein
CK305	fw-SingleCopy-OS	GACGTCTCACAGAAACGCGCTAGCCGCAGC	C. Klein
CK306	Rev-SingleCopy-OS	GACGTCTCACTACGAAGATCATCTTATTAATCAGATAAA ATATTTC	C. Klein
CK307	RBS-replace-rev	CACGTCTCTTGGATCTGATAAATTGTTAAAGAGCGTCGA AG	C. Klein
CK308	RBS7-replace-fw	GTCGTCTCTTCCAGCAGGAGCGATGAAATGGTGTCTATC ACTAAAGATC	C. Klein
CK309	RBS12-replace-fw	GTCGTCTCTTCCATCAGGAGGTGACAAATGGTGTCTATCA CTAAAGATC	C. Klein
CK310	17er-TG-lib-fw	GGTCTCGGACANNNNNNNNNNNNNNTGNTATACGAGA CCGTGTCTATCAC	C. Klein
CK311	17er-TGTG-lib-fw	GGTCTCGGACANNNNNNNNNNNNTGTGNTATACGAGA CCGTGTCTATCAC	C. Klein
CK312	17er-noT-lib-fw	GGTCTCGGACANNNNNNNNNNNNNNVNNTATACGAG ACCGTGTCTATCAC	C. Klein
CK313	17er-noG-lib-fw	GGTCTCGGACANNNNNNNNNNNNNNNHNTATACGAG ACCGTGTCTATCAC	C. Klein
CK314	fw-G8-referenz	GGAGTTGACATCTAGGTTAATGTGTGGTATAATTAGCTAC GAT	C. Klein
CK315	rev-G8-referenz	AGTAATCGTAGCTAATTATACCACACATTAACCTAGATG TCAA	C. Klein
CK316	fw-F1-referenz	GGAGTTGACACTCGAGGGAGGTCGTGTTATAATTAGCTA CGAT	C. Klein
CK317	rev-F1-referenz	AGTAATCGTAGCTAATTATAACACGACCTCCCTCGAGTG TCAA	C. Klein
CK318	rev-BBwoLead	GCCGTCTCACCATTTCATCGCTCCTGCTGG	C. Klein
CK319	rev-BBwLead	GCCGTCTCACCATTTGATCTTTAGTGATAGACACCAT	C. Klein
CK320	fw-BB	GCCGTCTCAGCTTCGGTTGCGCTCTCGTGC	C. Klein
CK321	fw-mVenusAT	GCCGTCTCAATGGTATCTAAAGGAGAAGAACTG	C. Klein
CK322	rev-mVenusAT	GCCGTCTCAAAGCATAAAACGAAAGGCCCAGTCTTTCGA CTGAGCCTTTCGTTTTATAAGCTTATTTATACAGTTCATCC ATTCC	C. Klein
CK323	fw-EMSA-OSBB-cy3	CTCTGGCGAAGACATGGAG	C. Klein

4.11 List of plasmids

Table 4.12: plasmids used in this work

ID	Name	Created by/Source	Antibiotic marker
P1	pICH41295	Weber et al 2011	
P2	pICH41308	Weber et al 2011	
P3	pICH41276	Weber et al 2011	
P4	pICH47732	Weber et al 2011	
P5	pICH47742	Weber et al 2011	
P6	pICH47751	Weber et al 2011	
P7	pICH47761	Weber et al 2011	
P8	pICH47802	Weber et al 2011	
P9	pICH47811	Weber et al 2011	
P10	pICH47822	Weber et al 2011	
P11	pICH47831	Weber et al 2011	
P12	pICH41780	Weber et al 2011	
P13	pICH41766	Weber et al 2011	
P14	pICH41744	Weber et al 2011	
P15	pICH89921	Weber et al 2011	
P16	pAGM4673	Weber et al 2011	
P17	pICH83966	Weber et al 2011	
P18	pMA60	Schindler et al. 2016	
P19	pMA61	Schindler et al. 2016	
P20	pMA62	Schindler et al. 2016	
P21	pMA63	Schindler et al. 2016	
P22	pMA64	Schindler et al. 2016	
P23	pMA65	Schindler et al. 2016	
P24	pMA66	Schindler et al. 2016	
P25	pICH50892	Schindler et al. 2016	
P26	pICH50872	Schindler et al. 2016	
P27	pICH50881	Schindler et al. 2016	
P28	pICH50927	Schindler et al. 2016	
P29	pICH50900	Schindler et al. 2016	
P30	pICH50914	Schindler et al. 2016	
P31	pICH50932	Schindler et al. 2016	
P34	pCK7	Carlo Klein	
P35	pCK2	Carlo Klein	
P36	pCK4	Carlo Klein	
P37	pCK31	Carlo Klein	
P38	pCK5	Carlo Klein	
P39	pCK9	Carlo Klein	
P40	pCK10	Carlo Klein	
P41	pCK11	Carlo Klein	
P42	pCK12	Carlo Klein	
P43	pCK13	Carlo Klein	
P44	pCK14	Carlo Klein	
P45	pCK15	Carlo Klein	
P46	pCK16	Carlo Klein	
P47	pCK17	Carlo Klein	

Continued on next page

ID	Name	Created by/Source	Antibiotic marker
P48	pCK18	Carlo Klein	
P49	pCK19	Carlo Klein	
P50	pCK20	Carlo Klein	
P51	pCK21	Carlo Klein	
P52	pCK22	Carlo Klein	
P53	pCK23	Carlo Klein	
P54	pCK24	Carlo Klein	
P55	pCK25	Carlo Klein	
P56	pCK26	Carlo Klein	
P57	pCK27	Carlo Klein	
P58	MCF-liBC	Carlo Klein	spec
P59	MCF-liBE	Carlo Klein	spec
P60	MCF-liBG	Carlo Klein	spec
P61	MCF-liFA	Carlo Klein	spec
P62	MCF-liHA	Carlo Klein	spec
P63	MCF-liIA	Carlo Klein	spec
P64	MCF-liDE	Carlo Klein	spec
P65	MCF-liDG	Carlo Klein	spec
P66	MCF-liDI	Carlo Klein	spec
P67	MCF-liFG	Carlo Klein	spec
P68	MCF-liFH	Carlo Klein	spec
P69	MCF-liJE	Carlo Klein	spec
P70	MCF-liHI	Carlo Klein	spec
P71	MCF-liHJ	Carlo Klein	spec
P72	MCF-liCA	Carlo Klein	spec
P73	MCF-liIG	Carlo Klein	spec
P74	MCF-posAB	Carlo Klein	kan
P75	MCF-posCD	Carlo Klein	kan
P76	MCF-posEF	Carlo Klein	kan
P77	MCF-posGH	Carlo Klein	kan
P78	MCF-posIJ	Carlo Klein	kan
P79	MCF-posAB-p15A	Carlo Klein	kan
P80	MCF-posCD-cmR	Carlo Klein	kann,cm
P81	MCF-posEF-mC	Carlo Klein	kan
P82	MCF-posGH-mV	Carlo Klein	kan
P83	MCF-posIJ-mTq2	Carlo Klein	kan
P93	pICH47732-LEVI 1	Patrick Sobetzko	amp
P94	pICH47732-LEVI-ssrA-vf 4	Patrick Sobetzko	amp
P95	pICH47732-LEVI-ssrA-mf 7	Patrick Sobetzko	amp
P108	LE1-gyrAp-pICH41233	Leonie Emde	Streptomycin
P109	LE2-gyrBp-pICH41233	Leonie Emde	Streptomycin
P110	LE3-topAp-pICH41233	Leonie Emde	Streptomycin
P111	LE4-dnaK-pICH41233	Leonie Emde	Streptomycin
P112	pICH41246-RBS-enhanced-leader-peptide	Leonie Emde	Streptomycin
P140	pICH41233-lacUV5-lexA408	Patrick Sobetzko	str,spec
P141	MCF-LinkBD	Carlo Klein	spec
P142	MCF-LinkBF	Carlo Klein	spec
P143	MCF-LinkBH	Carlo Klein	spec

Continued on next page

ID	Name	Created by/Source	Antibiotic marker
P144	MCF-LinkBI	Carlo Klein	spec
P145	MCF-LinkBJ	Carlo Klein	spec
P146	MCF-LinkAD	Carlo Klein	spec
P147	MCF-LinkAE	Carlo Klein	spec
P148	MCF-LinkAG	Carlo Klein	spec
P149	MCF-LinkAJ	Carlo Klein	spec
P150	MCF-LinkDF	Carlo Klein	spec
P151	MCF-LinkDH	Carlo Klein	spec
P152	MCF-LinkDJ	Carlo Klein	spec
P153	MCF-LinkCE	Carlo Klein	spec
P154	MCF-LinkCF	Carlo Klein	spec
P155	MCF-LinkCG	Carlo Klein	spec
P156	MCF-LinkCH	Carlo Klein	spec
P157	MCF-LinkCI	Carlo Klein	spec
P158	MCF-LinkCJ	Carlo Klein	spec
P159	MCF-LinkFI	Carlo Klein	spec
P160	MCF-LinkFJ	Carlo Klein	spec
P161	MCF-LinkEG	Carlo Klein	spec
P162	MCF-LinkEH	Carlo Klein	spec
P163	MCF-LinkEI	Carlo Klein	spec
P164	MCF-LinkGJ	Carlo Klein	spec
P183	pLE29-ribFp-pICH41233	Leonie Emde	str
P184	pLE30-pssAp-pICH41233	Leonie Emde	str
P185	pLE13-gyrAp-RBSenh-mV-T1-posGH	Leonie Emde	kan
P186	pLE14-gyrBp-RBSenh-mV-T1-posGH	Leonie Emde	kan
P187	pLE15-topAp-RBSenh-mV-T1-posGH	Leonie Emde	kan
P188	pLE16-dnaKp-RBSenh-mV-T1-posGH	Leonie Emde	kan
P189	pLE31-ribFp-RBSenh-mV-T1-posGH	Leonie Emde	kan
P190	pLE32-pssAp-RBSenh-mV-T1-posGH	Leonie Emde	kan
P191	pLE41-gyrAp-mV-MCFbb	Leonie Emde	cm
P192	pLE42-gyrBp-mV-MCFbb	Leonie Emde	cm
P193	pLE43-topAp-mV-MCFbb	Leonie Emde	cm
P194	pLE44-dnaKp-mV-MCFbb	Leonie Emde	cm
P195	pLE45-ribFp-mV-MCFbb	Leonie Emde	cm
P196	pLE46-pssAp-mV-MCFbb	Leonie Emde	cm
P197	pLE17-gyrBp-mC-mV-mTq-MCFbb-konvergent	Leonie Emde	cm
P198	pLE18-gyrBp-mC-mV-mTq-MCFbb-divergent	Leonie Emde	cm
P199	pLE51-gyrBp-mC-mV-mTq-MCFbb-tandem	Leonie Emde	cm
P200	pLE52-gyrBp-RBSenh-mC-T1-PosEF	Leonie Emde	kan
P201	pLE53-gyrBp-RBSenh-mTq2-T1-PosIJ	Leonie Emde	kan
P202	MCF-End-Linker-XA	Carlo Klein	spec
P203	MCF-End-Linker-XB	Carlo Klein	spec
P204	MCF-End-Linker-XC	Carlo Klein	spec
P205	MCF-End-Linker-XD	Carlo Klein	spec
P206	MCF-End-Linker-XE	Carlo Klein	spec
P207	MCF-End-Linker-XF	Carlo Klein	spec

Continued on next page

ID	Name	Created by/Source	Antibiotic marker
P208	MCF-End-Linker-XG	Carlo Klein	spec
P209	MCF-End-Linker-XH	Carlo Klein	spec
P210	MCF-End-Linker-XI	Carlo Klein	spec
P211	MCF-End-Linker-XJ	Carlo Klein	spec
P212	MCF-End-Linker-AY	Carlo Klein	spec
P213	MCF-End-Linker-BY	Carlo Klein	spec
P214	MCF-End-Linker-CY	Carlo Klein	spec
P215	MCF-End-Linker-DY	Carlo Klein	spec
P216	MCF-End-Linker-EY	Carlo Klein	spec
P217	MCF-End-Linker-FY	Carlo Klein	spec
P218	MCF-End-Linker-GY	Carlo Klein	spec
P219	MCF-End-Linker-HY	Carlo Klein	spec
P220	MCF-End-Linker-IY	Carlo Klein	spec
P221	MCF-End-Linker-JY	Carlo Klein	spec
P222	MCF-Destination	Carlo Klein	cm
P223	pNovoScreen (Spacer Backbone)	Carlo Klein	cm
P229	BackBone OptoScreen fw	Carl Weile	cm1/2
P230	BackBone OptoScreen rev	Carl Weile	cm1/2
P245	BackBone OptoScreen fw T2	Carl Weile	cm1/2
P246	BackBone OptoScreen rev T2	Carl Weile	cm1/2
P261	OS-bb-rev-noEsp3I	Carlo Klein	cm 1/2
P262	OS-bb-rev-noEsp3I-RBS7	Carlo Klein	cm 1/2
S263	OS-bb-rev-noEsp3I-RBS12	Carlo Klein	cm 1/2

4.12 Source code

4.12.1 NovoScreen.v1

```
1
2  ######################## spacer analysis novoscreen  ################
3
4  ### Packages
5
6
7  library(stringr)
8  library(tcltk)
9  library(gdata)
10 library(ggplot2)
11 library(ggseqlogo) # https://omarwagih.github.io/ggseqlogo/
12 library(seqinr)
13 library(randomForest)
14
15
16
17 ### Importing
18
19
20 folder <- tk_choose.dir(getwd(), "Choose an experiment folder containing the .exls files") # this command returns
       the folder
21 files <- list.files(folder, full.names = T) # creates a variable with all filenames of the folder
```

```
Data.files.xls <- grep("Annotation", files, invert = T)   # same as for annofiles but greps all files that don not contain annotation
Data.files.xls <- sort(files[Data.files.xls])


spacer <- 10:23 ### 10 = no -35 and 11 = N-only controls 12:23 spacers

PlateRows <- c("A","B","C","D","E","F","G","H")
posneg.ratio <- c()
mV.list <- list()
mV.treat.list <- list()
mV.ratio.list <- list()

mC.linear <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
mC.novo.linear <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
mV.linear <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
mV.novo.linear <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
OD.linear <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
OD.novo.linear <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
mV.linear.anno <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
OD.dif.linear <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
eleminated <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())

for(i in 1:length(Data.files.xls))
{
  print(Data.files.xls[i])
  OD <- read.xls(Data.files.xls[i], sheet=1, skip=19, nrows=8)
  OD <- OD[,-1]
  col.no.spacer <- read.xls(Data.files.xls[i], sheet=1, skip=18, nrows=1)
  col.no.spacer <- col.no.spacer[,-1]
  mV <- read.xls(Data.files.xls[i], sheet=1, skip=42, nrows=8)
  mV <- mV[,-1]
  mC <- read.xls(Data.files.xls[i], sheet=1, skip=65, nrows=8)
  mC <- mC[,-1]
  OD.novo <- read.xls(Data.files.xls[i], sheet=2, skip=19, nrows=8)
  OD.novo <- OD.novo[,-1]
  mV.novo <- read.xls(Data.files.xls[i], sheet=2, skip=42, nrows=8)
  mV.novo <- mV.novo[,-1]
  mC.novo <- read.xls(Data.files.xls[i], sheet=2, skip=65, nrows=8)
  mC.novo <- mC.novo[,-1]

  OD.blank <- OD-0.045
  mC.blank <- mC-5
  mV.blank <- mV-40
  OD.novo.blank <- OD.novo-0.045
  mC.novo.blank <- mC.novo-5
  mV.novo.blank <- mV.novo-40

  mC.norm <- mC.blank/OD.blank
  mC.novo.norm <- mC.novo.blank/OD.novo.blank
  mV.norm <- mV.blank/OD.blank
  mV.novo.norm <- mV.novo.blank/OD.novo.blank
  OD.dif <- OD.blank - OD.novo.blank


  for(row in 1:8)
  {
    for(col in 1:12)
    {
      if(mC.norm[row,col] >= 700 & mC.norm[row,col] <= 2000 & abs(OD.dif[row,col]) <= 0.05 & OD.blank[row,col] > 0.5 & OD.novo[row,col] > 0.5 & mC.novo.norm[row,col] >= 700 & mC.novo.norm[row,col] <= 2000)
        {
      mC.linear[[col.no.spacer[1,col]]] <- append(mC.linear[[col.no.spacer[1,col]]],max(0,mC.norm[row,col]));
```

```
        mC.novo.linear[[col.no.spacer[1,col]]] <- append(mC.novo.linear[[col.no.spacer[1,col]]],max(0,mC.novo.norm[row,col]));
        mV.linear[[col.no.spacer[1,col]]] <- append(mV.linear[[col.no.spacer[1,col]]],max(0,mV.norm[row,col]));
        mV.linear.anno[[col.no.spacer[1,col]]] <- append(mV.linear.anno[[col.no.spacer[1,col]]], paste(sub('D:/Carlo/Doktorarbeit/Projects/SupercoilingSensitivePromoters/RelaxationSensitive(Novo)/DataRAW/','',Data.files.xls[i],fixed = T),PlateRows[row],col, sep = "_"));
        mV.novo.linear[[col.no.spacer[1,col]]] <- append(mV.novo.linear[[col.no.spacer[1,col]]],max(0,mV.novo.norm[row,col]));
        OD.dif.linear[[col.no.spacer[1,col]]] <- append(OD.dif.linear[[col.no.spacer[1,col]]],OD.dif[row,col]);
        OD.linear[[col.no.spacer[1,col]]] <- append(OD.linear[[col.no.spacer[1,col]]],max(0,OD.blank[row,col]));
        OD.novo.linear[[col.no.spacer[1,col]]] <- append(OD.novo.linear[[col.no.spacer[1,col]]],max(0,OD.novo.blank[row,col]));
          }
      }
    }
}


mean.ratio <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
mC.ratio <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
mV.ratio <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
mV.ratio.anno <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
mV.max.exp <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
mV.OD.dif <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
spacer.length <- list(c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c(),c())
clones <- c()

for(k in 1:23)
{
  if(length(mC.linear[[k]]) > 0)
    {
    print(k)
  mC.ratio[[k]] <- log(mC.novo.linear[[k]]/mC.linear[[k]],2)
  mV.ratio[[k]] <- log(mV.novo.linear[[k]]/mV.linear[[k]],2)
  mV.max.exp[[k]] <- mV.linear[[k]]
  mV.ratio.anno[[k]] <- mV.linear.anno[[k]]
  spacer.length[[k]] <-k
  mV.OD.dif[[k]] <- OD.dif.linear[[k]]
  mV.ratio[[k]][which(!is.finite(mV.ratio[[k]]))] <- NA
  #mV.ratio.anno[[k]][which(!is.finite(mV.ratio[[k]]))] <- NA
  mV.max.exp[[k]][which(!is.finite(mV.ratio[[k]]))] <- NA
  mV.OD.dif[[k]][which(!is.finite(mV.ratio[[k]]))] <- NA


  mV.ord.last <- order(mV.ratio[[k]])
  mV.ord.first <- order(mV.ratio[[k]], na.last=FALSE)
  mV.ord.last.pos <- order(abs(mV.ratio[[k]]))
  posneg.ratio <- append(posneg.ratio, length(which(mV.ratio[[k]] >= 0))/length(which(mV.ratio[[k]] < 0)))
  mean.ratio <- append(mean.ratio, sum(mV.ratio[[k]])/length(mV.ratio[[k]]))


  color <- c(rep("#AAAAFFFF", length(which(mV.ratio[[k]]<0))),rep("#FF4444FF", length(which(!(mV.ratio[[k]]<0)))))
  barplot(mV.ratio[[k]][mV.ord.last], main=paste(k, " bp ","spacer ", sep=""), col=color, ylim=c(-1.5,0.5))
  mtext(paste('N = ', length(mV.ratio[[k]][mV.ord.last])),side=4)
    }
}
### Selection of particular clones ###

clones <- do.call(rbind,mapply(cbind, mV.ratio.anno, mV.ratio, mV.max.exp, mV.OD.dif, spacer.length))
clones <- as.data.frame(clones, stringsAsFactors = F)

clones <- clones[complete.cases(clones),]

clones[,2] <- as.numeric(as.character(clones[,2]))
clones[,3] <- as.numeric(as.character(clones[,3]))
clones[,5] <- as.numeric(as.character(clones[,5]))

```

```
clones[which(clones[,5] <= 15),6] <- "supoptimal spacing"
clones[which(clones[,5] == 16),6] <- "optimal spacing"
clones[which(clones[,5] == 17),6] <- "optimal spacing"
clones[which(clones[,5] == 18),6] <- "optimal spacing"
clones[which(clones[,5] == 19),6] <- "optimal spacing"
clones[which(clones[,5] >= 20),6] <- "supoptimal spacing"

names(clones) <- c("Clone", "mV.Ratio","max.expression","OD.difference","Spacer Length", "Spacer Type")


clones[,1] <- sub(folder,'', clones[,1])
clones[,1] <- sub('.xlsx','', clones[,1])
clones[,1] <- substring(clones[,1], 20)

### for sequencing and stuff

clones.sub <- clones[grep(17, clones[,5]),]

clones.seq <- clones.sub[which(clones.sub[,3] >= 8000),]

clones.seq.sub <- clones.seq[order(clones.seq$max.expression),]

clones.seq.sub.sub <- clones.seq.sub[c(1:48,(nrow(clones.seq.sub)-47):nrow(clones.seq.sub)),]

### write table order plate names because easy pipetting

clones.seq.sub.sub <- clones.seq.sub.sub[order(clones.seq.sub.sub[,1]),]

write.table(clones.seq.sub.sub, file=file.path(choose.dir(),paste('spacer_',clones.seq.sub.sub[1,5],'_for_
    sequencing.txt')), row.names = F, col.names = T )

pos.responder <- clones.sub.sub[which(clones.sub.sub[,2] > 0.1),]
no.responder  <- clones.sub.sub[which(clones.sub.sub[,2] >= -0.1 & clones.sub[,2] <= 0.1),]
neg.responder <- clones.sub.sub[which(clones.sub.sub[,2] < -0.1),]

pos.responder <- pos.responder[order(pos.responder$mV.Ratio, pos.responder$max.expression),]
no.responder <- no.responder[order(no.responder$mV.Ratio, no.responder$max.expression),]
neg.responder <- neg.responder[order(neg.responder$mV.Ratio, neg.responder$max.expression),]


write.table(pos.responder, file=file.path(choose.dir(),paste('spacer_',clones.sub.sub[1,5],'_pos.responder.txt')),
    row.names = F, col.names = T )
write.table(no.responder, file=file.path(choose.dir(),paste('spacer_',clones.sub.sub[1,5],'_no.responder.txt')),
    row.names = F, col.names = T )
write.table(neg.responder, file=file.path(choose.dir(),paste('spacer_',clones.sub.sub[1,5],'_neg.responder.txt')),
    row.names = F, col.names = T )


### generating the plots ###

##Ratio vs Expression

p <- ggplot(clones.sub,aes(clones.sub$max.expression, clones.sub$mV.Ratio)) + geom_point(size=2, color = "
    cornflowerblue")
p <- p + stat_smooth(method = "lm", col = "red")
p <- p + theme_bw()
p

linearMod <- lm(mV.Ratio ~ max.expression, data=clones)  # build linear regression model on full data
print(linearMod)
#plot(linearMod)
lm.var <- coefficients(linearMod)

mv.ratio <- clones$mV.Ratio
expression <- clones$max.expression
cor <- cor(mv.ratio,expression) #correlation coefficient

clones.sub.corrected <- clones

```

```
clones.sub.corrected$mV.Ratio <- clones.sub.corrected$mV.Ratio - (clones.sub.corrected$max.expression*lm.var[2] + lm.var[1])


p <- ggplot(clones.sub.corrected,aes(clones.sub.corrected$max.expression, clones.sub.corrected$mV.Ratio, color = clones.sub.corrected$`Spacer Type`)) + geom_point(size=2)+
  scale_color_manual(values = c("#AAAAFFFF","#FF4444FF"))

p <- p + theme_bw()
p <- p + theme(legend.position="bottom")
p


p <- ggplot(data = clones.sub.corrected,aes(clones.sub.corrected[,5],clones.sub.corrected[,2])) + #colour = clones[,6])) +
  geom_jitter(position=position_jitter(0.2), show.legend = F, size = sqrt(clones.sub.corrected[,3]/1500), alpha=0.2) +
  scale_x_continuous(name = "spacer length [bp]", breaks = c(10,12,12,13,14,15,16,17,18,19,20,21,22,23)) +
  scale_y_continuous(name = "treated/untreated ratio (log2)")
p <- p + stat_summary(fun.y = median, geom="point", shape=18,
                      size=3, color="red") + theme_bw()
p <- p + theme(legend.position="bottom")
p


p <- ggplot(clones,aes(clones$max.expression, clones$mV.Ratio, color = clones$`Spacer Type`)) + geom_point(size=2)+
  scale_color_manual(values = c("#AAAAFFFF","#FF4444FF"))
p <- p + stat_smooth(method = "lm", color = "black")
p <- p + theme_bw()
p <- p + theme(legend.position="bottom")
p


p <- ggplot(clones.seq,aes(clones.seq$max.expression, clones.seq$mV.Ratio)) + geom_point(size=2, color = "cornflowerblue")+
    geom_vline(xintercept = clones.seq.sub[48,3],colour="#BB0000", linetype="dashed")+
    geom_vline(xintercept = clones.seq.sub[(nrow(clones.seq.sub)-47),3],colour="#BB0000", linetype="dashed")

p <- p + theme_bw(base_size = 22)

p
## jitter plot mV.ratios ##


p <- ggplot(data = clones,aes(clones[,5],clones[,2])) + #colour = clones[,6])) +
  geom_jitter(position=position_jitter(0.25), show.legend = F, size = sqrt(clones[,3]/1500), alpha=0.2) +
  scale_x_continuous(name = "spacer length [bp]", breaks = c(10,11,12,13,14,15,16,17,18,19,20,21,22,23)) +
  scale_y_continuous(name = "treated/untreated ratio (log2)")
p <- p + stat_summary(fun.y = median, geom="point", shape=18,
                  size=3, color="red") + theme_bw()

p <- p + theme(legend.position="bottom")
p

ggsave( "plot1.tiff", p, dpi = 300)
##

barplot(log(posneg.ratio), main="activated/repressed (log2)", names = 10:23, ylim = c(-5,5))#, xlab="spacer length")

boxplot(mV.linear, main="expression level", names = 1:24, xlab="spacer length", col="#CCCCFFFF",ylim = c(0, 40000), xlim = c(12,23))

p <- ggplot(data=clones, aes(x=clones[,5],y=clones[,3],group = 1))+ scale_y_log10()  +
  scale_x_continuous(name = "spacer length [bp]", breaks = c(11,12,13,14,15,16,17,18,19,20,21,22,23))+
  geom_boxplot(aes(group = cut_width(clones[,5], 1)),outlier.alpha = 0, fill="orange")+
  geom_jitter(position=position_jitter(0.2), show.legend = F, alpha=0.2)+
  geom_vline(xintercept = 11.5,colour="grey", linetype="dashed")

```

```
p <- p + theme_bw()

p

ggsave("plot2.pdf", plot= last.plot(),device = "pdf", dpi = 600)

boxplot(mV.novo.linear, main="expression level novobiocin", names = 1:24, xlab="spacer length", col="#FFCCCCFF",
    ylim = c(0, 40000),xlim = c(12,23))


### scatter OD.dif and Ratio

order <- order(as.numeric(clones.sub[,4]))
p <- ggplot(clones.sub) + geom_point(aes(clones$OD.difference[order], clones$mV.Ratio[order], color = as.numeric(
    clones[order,5])))
p +  theme_Publication()

  xlab('OD difference') + ylab('mV ratio')

# oder

plot(clones$max.expression,clones$mV.Ratio,xlab='OD difference',ylab='mV ratio',
     col = ifelse(clones$max.expression > median(clones$max.expression),'red','green'),
     pch = 19, mtext(paste('median = ', median(clones$max.expression)), side = 3, adj = 1))
legend("topright", inset=.05,
       c("> median","< median"),  fill = c("red" , "green"), horiz=F)
```

4.12.2 PromoterSequencingAnalysis_v1

```

############## sequencing analysis ################

library(tcltk)
library(gdata)
library(ggplot2)
library(gridExtra)
library(reshape2)
library(ggseqlogo)
library(stringr)
library(stringdist)
library(ggseqlogo) # https://omarwagih.github.io/ggseqlogo/
library(seqinr)


#### import FASTAs ####

folder.fasta <- tk_choose.dir(getwd(), "Choose a suitable folder") # this command returns the folder
files.fasta <- list.files(folder.fasta, full.names = T)
Data.files.fasta <- grep(".fasta", files.fasta)
Data.files.fasta <- sort(files.fasta[Data.files.fasta])

file=file.choose()
bend.table <- read.csv(file,sep = ";",stringsAsFactors = FALSE)[,c(1,2,3)]
#bend.table <- as.data.frame(bend.table[2:nrow(bend.table),])
bend.table <- bend.table[2:nrow(bend.table),]
bend.table[,2] <- as.numeric(bend.table[,2])
bend.table[,3] <- as.numeric(bend.table[,3])


file.melting=file.choose()
melt.table <- read.csv(file.melting,sep = ";",stringsAsFactors = FALSE)[2:17,c(1,2,3,4)]
#bend.table <- as.data.frame(bend.table[2:nrow(bend.table),])
melt.table[,2] <- as.numeric(melt.table[,2])
melt.table[,3] <- as.numeric(melt.table[,3])
melt.table[,4] <- as.numeric(melt.table[,4])

```

```
38  sequences.fasta.df <- c()
39
40  for(i in 1:length(Data.files.fasta))
41  {
42    print(Data.files.fasta[i])
43
44    sequences.fasta <- read.fasta(file = Data.files.fasta[i],
45                                  seqtype = c("DNA"), as.string = TRUE, forceDNAtolower = FALSE)#,
46    # set.attributes = FALSE, legacy.mode = TRUE, seqonly = TRUE, strip.desc = FALSE,
47    # bfa = FALSE, sizeof.longlong = .Machine$sizeof.longlong,
48    # endian = .Platform$endian, apply.mask = TRUE)
49
50    sequences.fasta.df <- rbind(sequences.fasta.df,sequences.fasta)
51  }
52
53  sequences.fasta.df <- as.data.frame(sequences.fasta.df)
54  sequences.fasta.df[,2] <- sequences.fasta.df
55  sequences.fasta.df[,1] <- Data.files.fasta
56
57  sequences.fasta.df[,1] <- sub(folder.fasta,'', sequences.fasta.df[,1])
58  sequences.fasta.df[,1] <- substring(sequences.fasta.df[,1], 2,4)
59  sequences.fasta.df[,1] <- sub("_",'', sequences.fasta.df[,1])
60
61  seqs <- sequences.fasta.df
62  pos35 <- regexpr('TTGACA',seqs[,2]);
63  pos10 <- regexpr('TATAAT',seqs[,2]);
64  spacer.seq <- c();
65  discriminator <- c();
66  for(i in 1:nrow(seqs))
67  {
68    if(pos10[i] > 0 & pos35[i] > 0)
69    {
70      spacer.seq <- rbind(spacer.seq, unlist(c(seqs[i,1:2],substr(seqs[i,2], pos35[i]+4, pos10[i]+1))));
71      discriminator <- rbind(discriminator, unlist(c(substr(seqs[i,2], pos10[i], pos10[i]+27))));
72    }
73  }
74
75
76  spacer.seq <- cbind(spacer.seq, nchar(spacer.seq[,3]))
77  gc <- c()
78
79  for(i in 1:nrow(spacer.seq))
80  {
81    seq <- strsplit(unlist(spacer.seq[i,3]), "")[[1]];
82    gc <- c(gc, length(which(seq == "G" | seq == "C"))/nchar(spacer.seq[i, 3]));
83  }
84
85  spacer.seq <- cbind(spacer.seq,gc);
86  spacer.seq <- as.data.frame(spacer.seq)
87  spacer.seq <- spacer.seq[complete.cases(spacer.seq),]
88  spacer.seq <- spacer.seq[which(spacer.seq$V4 == 21),]
89  #spacer.seq[,1] <- as.character(spacer.seq[,1])
90  spacer.seq[,2] <- as.character(spacer.seq[,2])
91  spacer.seq[,3] <- as.character(spacer.seq[,3])
92  spacer.seq$V4 <- 21
93
94
95
96  rf.data.line <-c()
97  for(i in 1:nrow(spacer.seq))
98  {
99    seq <- strsplit(spacer.seq[i,3], "")[[1]]
100   print(paste(i,seq, sep = "␣"))
101   bendability.curvature <- c()
102   bendability <- c()
103   melting.temp <- c()
104   for(j in 1:length(seq)) ### fehler!!!! nimmt den level nicht den bendability wert -> solved see above import of
          bend.table
105   {
```

```
    if(j<=length(seq)-2)
    {
      bendability.curvature  <- c(bendability.curvature, bend.table[which(bend.table[,1] == paste(seq[j],seq[j+1],seq[j+2],sep = "")),2])
      bendability  <- c(bendability, bend.table[which(bend.table[,1] == paste(seq[j],seq[j+1],seq[j+2],sep = "")),3])
    }
    if(j <= length(seq)-1 )
    {
      melting.temp <- c(melting.temp, melt.table[which(melt.table[,1] == paste(seq[j],seq[j+1],sep = "")),4])
    }

  }


  spacer.seq[i,6] <- mean(bendability.curvature)
  spacer.seq[i,7] <- mean(bendability)
  spacer.seq[i,8] <- mean(melting.temp)


  #rf.data.line <- rbind(rf.data.line,cbind(unlist(spacer.seq[i,]),t(as.data.frame(bendability)),t(as.data.frame(melting.temp))))
  print(paste(i,length(unlist(spacer.seq[i,])),length(bendability),length(melting.temp), sep = "␣"))
  rf.data.line <- rbind(rf.data.line,c(unlist(spacer.seq[i,]),bendability,melting.temp))

}

rf.data.line <- as.data.frame(rf.data.line)
rf.data.line[,1] <- spacer.seq[,1]

for(i in 1:nrow(rf.data.line))
{
  rf.data.line[i,48] <- paste(grep(rf.data.line[i,3],rf.data.line[,3]), collapse = ";")
}


names(rf.data.line) <- c("Well","Sequence","Spacer","length",
                         "GC-content","Bendability.Curvature","Bendability",
                         "Melting.Energy","bend1","bend2","bend3","bend4","bend5",
                         "bend6","bend7","bend8","bend9","bend10","bend11","bend12",
                         "bend13","bend14","bend15","bend16","bend17","bend18","bend19",
                         "melt1","melt2","melt3","melt4","melt5","melt6","melt7","melt8",
                         "melt9","melt10","melt11","melt12","melt13","melt14","melt15",
                         "melt16","melt17","melt18","melt19","melt20","Dublicates")


for(i in 1:nrow(spacer.seq))
{
  spacer.seq[i,9] <- paste(grep(spacer.seq[i,3],spacer.seq[,3]), collapse = ";")
}

names(spacer.seq) <- c("Well","Sequence","Spacer","length","GC-content","Bendability.Curvature","Bendability","Melting.Energy","Dublicates")

high.expressed <- c("F10","A4","F1","H8","E5","C6","C2","F3","B6","C10","E10","F6","E8","E12","D8","H7","E1","C1","D9","D4","E11","D5","C8","H9","B10","D11","A1","G12","A10","C5","D7","H1","G3","H10","C9","A5","A6","F4","A3","H3","F2","C7","C12","H6","B5","F8","F9","G7")

for.seq <- read.xls(file.choose())


df <- c()
df <- merge(rf.data.line, for.seq)

for(i in 1:nrow(df))
{
  if(df[i,1] %in% high.expressed)
    df[i,53] <- "right.group"
  else
```

```
    df[i,53] <- "left.group"
}

df <- df[unique(df$Dublicates),]
df <- df[,c(1:47,50,51,53)]
df <- df[complete.cases(df),]
df$`GC-content` <- as.numeric(df$`GC-content`)
df$Bendability.Curvature <- as.numeric(df$Bendability.Curvature)
df$Bendability <- as.numeric(df$Bendability)
df$Melting.Energy <- as.numeric(df$Melting.Energy)

p <- ggplot(data=df, aes(x=df$V53,y=df$`GC-content`, fill="cornflowerblue"))+#, color = df$`GC-content`))+
  #geom_point()
  #+
  geom_boxplot()+
  geom_jitter(position=position_jitter(0.2), show.legend = F, alpha=0.2)
#geom_vline(xintercept = 11.5,colour="grey", linetype="dashed")

p + theme_bw()

ggsave("GC.pdf", device = "pdf", dpi = 600, path = tk_choose.dir(getwd(), "Choose a suitable folder"))

### melting

df.splt.rg <- df[which(df$V53 == "right.group"),]
df.splt.lg <- df[which(df$V53 == "left.group"),]

melt.rg <- df.splt.rg[,c(29:46)]
melt.rg <- as.matrix(melt.rg)
melt.rg <- data.matrix(melt.rg)

mean.melt.rg <- apply(melt.rg, MARGIN = c(1,2), FUN = as.numeric)
mean.rg <- apply(mean.melt.rg, MARGIN = 2, FUN = mean)
sd.rg <- apply(mean.melt.rg, MARGIN = 2, FUN = sd)/sqrt(nrow(mean.melt.rg))

melt.lg <- df.splt.lg[,c(29:46)]

melt.lg <- as.matrix(melt.lg)
melt.lg <- data.matrix(melt.lg)

mean.melt.lg <- apply(melt.lg, MARGIN = c(1,2), FUN = as.numeric)
mean.lg <- apply(mean.melt.lg, MARGIN = 2, FUN = mean)
sd.lg <- apply(mean.melt.lg, MARGIN = 2, FUN = sd)/sqrt(nrow(mean.melt.lg))


p <- ggplot() +
  geom_line(aes(x=1:18, y=mean.rg, colour ="strong promoters"))+
  geom_ribbon(aes(x=1:18, ymin=(mean.rg-sd.rg), ymax=(mean.rg+sd.rg)),alpha=0.25)+
  geom_line(aes(x=1:18,y=mean.lg, colour ="weak promoters"))+
  geom_ribbon(aes(x=1:18, ymin=(mean.lg-sd.lg), ymax=(mean.lg+sd.lg)), alpha=0.25)

p  + xlab("spacer position [bp]") + ylab("melting engergy [kJ/mol]") + theme_bw()+theme(legend.position = "bottom")

### divide groups to compare bendability

df.splt.rg <- df[which(df$V53 == "right.group"),]
df.splt.lg <- df[which(df$V53 == "left.group"),]

bend.rg <- df.splt.rg[,c(9:27)]
bend.rg <- as.matrix(bend.rg)
bend.rg <- data.matrix(bend.rg)

mean.bend.rg <- apply(bend.rg, MARGIN = c(1,2), FUN = as.numeric)
mean.rg <- apply(mean.bend.rg, MARGIN = 2, FUN = mean)
sd.rg <- apply(mean.bend.rg, MARGIN = 2, FUN = sd)/sqrt(nrow(mean.bend.rg))

bend.lg <- df.splt.lg[,c(9:27)]

bend.lg <- as.matrix(bend.lg)
```

```
238 bend.lg <- data.matrix(bend.lg)
239
240 mean.bend.lg <- apply(bend.lg, MARGIN = c(1,2), FUN = as.numeric)
241 mean.lg <- apply(mean.bend.lg, MARGIN = 2, FUN = mean)
242 sd.lg <- apply(mean.bend.lg, MARGIN = 2, FUN = sd)/sqrt(nrow(mean.bend.lg))
243
244
245 p <- ggplot() +
246   geom_line(aes(x=1:19, y=mean.rg, color = "strong␣promoters"))+
247   geom_ribbon(aes(x=1:19, ymin=(mean.rg-sd.rg), ymax=(mean.rg+sd.rg) ),alpha=0.25) +
248   geom_line(aes(x=1:19,y=mean.lg, color ="weak␣promoters"))+
249   geom_ribbon(aes(x=1:19, ymin=(mean.lg-sd.lg), ymax=(mean.lg+sd.lg) ), alpha=0.25)
250
251 p  + xlab("spacer␣position␣[bp]") + ylab("bendability") + theme_bw()+theme(legend.position = "bottom")
```

List of Figures

Figures 1.1, 1.2, 1.3, 1.4, 2.6 and 2.14 are made with BioRender

List of Tables

Appendix

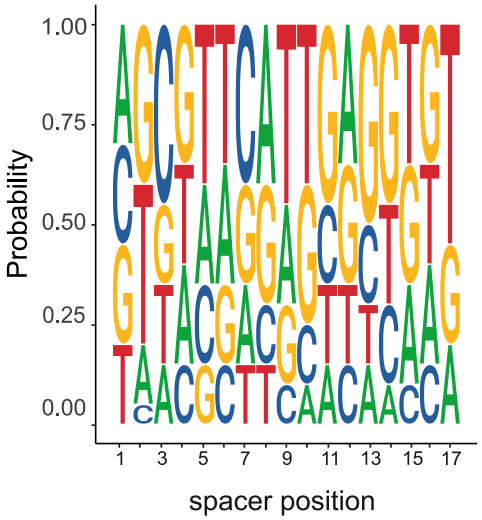


Figure 4.3: Sequence logo of the 17 bp spacer library oligonucleotides when tested in a plasmid without creating a promoter.

Table 4.13: Sequences, expression and mVenus ratio of the spacers used for the analysis in section 2.2.4

	Spacer	GC	Bend.	Melt.	mV.Ratio	max.expression	group
1	CATATCTAATGATCGGTGGTA	38.10	5.91	-5.16	-1.05	28957.09	right.group
3	CATTTCATGGTAGATCTGCTA	38.10	5.94	-5.24	-0.36	16166.17	left.group
4	CAAGAGCCGTGTGCTATGGTA	52.38	6.21	-6.03	-0.22	16751.12	left.group
5	CAGTCGAACAGATAGATGTTA	38.10	5.89	-5.24	-0.59	27143.32	right.group
6	CATCCCGGACAGGTCATGTTA	52.38	6.05	-5.92	-0.61	27829.14	right.group
7	CAGGACTCCTCCAGCGTGGTA	61.90	5.45	-6.32	-0.55	27803.28	right.group
8	CATGGACGGTTGCTTATGATA	42.86	6.22	-5.57	-0.43	17583.99	left.group
10	CAGGGAGCGGGTTCGATGGTA	61.90	5.27	-6.40	-0.72	29350.77	right.group
13	CATATCGTCTGCGGTGTGTTA	47.62	6.24	-5.85	-0.33	16267.74	left.group
14	CAGGAGGATCCATTTGTGGTA	47.62	4.84	-5.65	-0.40	12539.11	left.group
15	CAAGACCTACGTCATGTGGTA	47.62	6.05	-5.70	-0.42	16116.87	left.group
16	CACGTGGTTCTGCTCGTGATA	52.38	6.01	-6.09	-0.65	26874.69	right.group
17	CATTCGTGGATTGATATGGTA	38.10	5.38	-5.29	-0.16	14109.84	left.group
18	CATTAAGCGCGTCATGTGTTA	42.86	5.91	-5.78	-0.19	16264.29	left.group
21	CAATGCATCTCGGGTATGATA	42.86	6.48	-5.51	-0.22	17494.76	left.group
22	CATTTCCGGGTTACGATGGTA	47.62	4.63	-5.77	-0.55	26892.67	right.group
23	CACGCATACTAACAAATGGTA	38.10	5.28	-5.40	-0.72	30775.35	right.group
24	CACCAAACCTTTTTAATGGTA	33.33	3.72	-5.17	-0.36	17722.12	left.group
25	CACTCTGAGGCAAATGTGATA	42.86	6.28	-5.54	-0.20	14578.21	left.group
29	CAGTCTATCTGGACCGTGGTA	52.38	5.57	-5.81	-0.66	27837.06	right.group
30	CAATTCAGGGGCGGTGTGGTA	57.14	5.23	-6.25	-0.35	16902.17	left.group
31	CATCCTCCCTACTATGTGATA	42.86	6.56	-5.24	-0.10	16484.81	left.group
32	CAATTAAAGCTTCTTAAGGTA	28.57	4.88	-4.87	-0.65	29299.60	right.group
33	CACCAGAGCCTATATGTGGTA	47.62	6.58	-5.58	-0.16	13591.99	left.group
34	CAGGTCGTATGTAGCGTGATA	47.62	6.67	-5.73	0.62	8807.19	left.group
35	CATTATCGCGGTGATGTGCTA	47.62	6.06	-5.89	-0.21	15680.37	left.group
36	CAAGTGTACCTAGGGATGGTA	47.62	5.50	-5.53	-0.65	29691.61	right.group
37	CAACCCTTGCCGTTTAGTGTA	47.62	5.03	-5.85	-0.60	29562.61	right.group
38	CACCCTTGCCCTTTTGTGGTA	52.38	4.97	-6.04	-0.25	14480.54	left.group
42	CAAATACTAGATTTTCGGTTA	28.57	4.12	-4.86	-0.58	29627.35	right.group
45	CAAATTTGATCATGTGTGGTA	33.33	4.98	-5.21	-0.23	14742.02	left.group
46	CATGTCATCGGCTATGTGGTA	47.62	6.61	-5.75	-0.28	17005.30	left.group
48	CATTTAAACGGCTTCGTGATA	38.10	5.17	-5.48	-0.17	16425.67	left.group
49	CAGTTGCTCTTAAGTCTGTTA	38.10	5.67	-5.30	-0.71	30108.44	right.group
52	CAAACGATCAATGATGATGTA	33.33	5.67	-5.19	-0.74	32252.19	right.group
53	CATGCAAATCAAAACATGGTA	33.33	5.04	-5.31	-0.18	14455.15	left.group
54	CATCCGCGACCAAAAATGGTA	47.62	3.94	-5.95	-0.35	16122.32	left.group
60	CATTTACGAGGACATGTGGTA	42.86	5.51	-5.53	-0.20	15978.33	left.group
62	CAACTACGCGTCGAAATGCTA	47.62	5.04	-5.99	-0.71	26620.39	right.group
63	CAAGTCTTCTTGTTTGTGGTA	38.10	4.70	-5.38	-0.10	13587.05	left.group
70	CAGTCCTTACCGAATGTGTTA	42.86	5.32	-5.52	-0.29	17595.02	left.group
72	CAGATATATTTCTCCGTGATA	33.33	6.38	-4.93	-0.56	26391.64	right.group
73	CATCTAGGTTAATGTGTGGTA	38.10	5.62	-5.19	-0.10	13312.25	left.group
76	CAAATCTTCTGTATTCGTGTA	33.33	5.20	-5.12	-0.51	28118.30	right.group
81	CAACTGTGCGGGAATGTGCTA	52.38	5.62	-6.15	-0.17	13394.54	left.group
82	CAAGATCTGGGGTTTGTGCTA	47.62	5.24	-5.76	-0.28	16750.37	left.group
83	CATGTCAACGACGCCGTGCTA	57.14	6.18	-6.46	-0.63	26884.32	right.group
85	CAATTATCAATTACGGTGGTA	33.33	4.69	-5.10	-0.69	31418.44	right.group
86	CAATTTAGGGATCGGATGGTA	42.86	4.46	-5.44	-0.56	29410.64	right.group

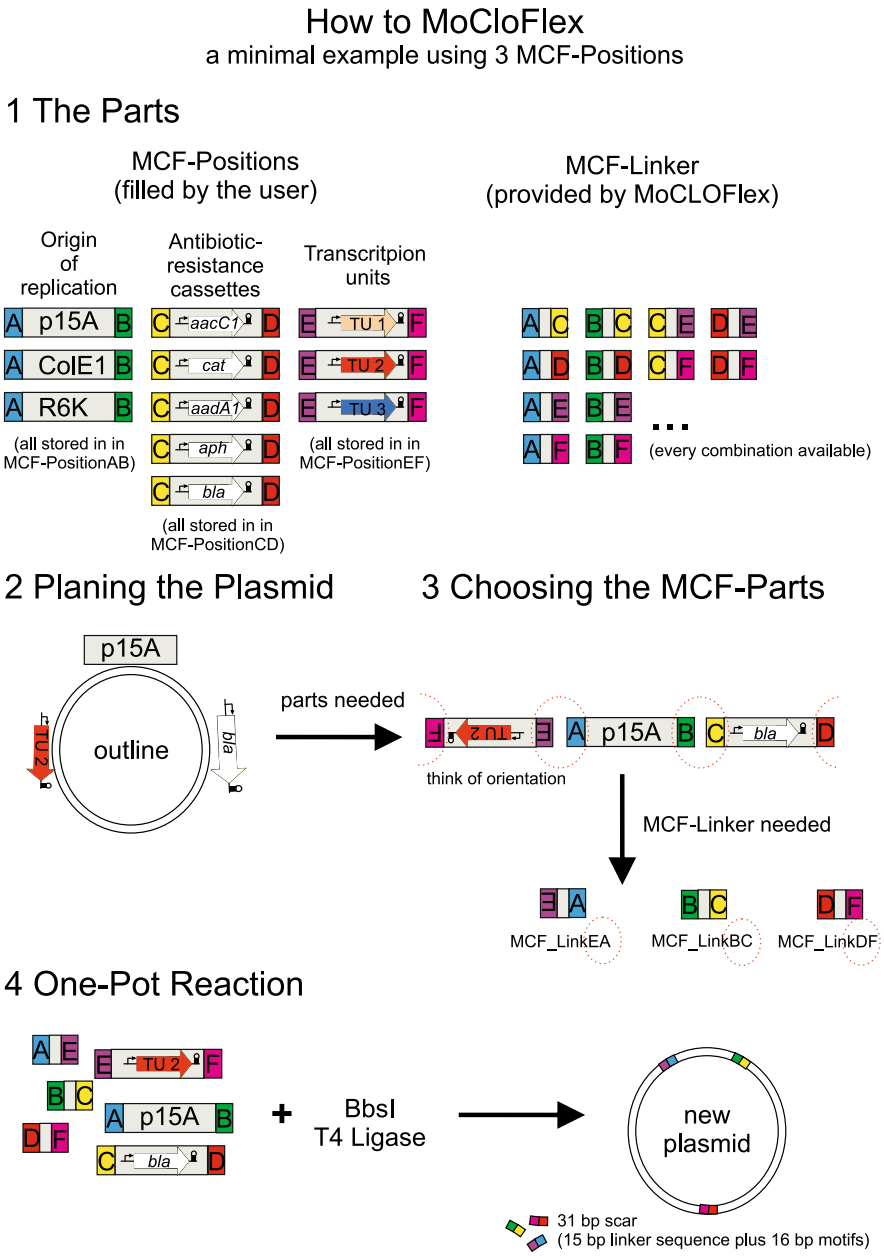


Figure 4.4: In this example 3 of 5 possible positions are used. In MCF-PositionAB different ori, in MCF- PositionCD antibiotic resistance cassettes, and in MCF-PositonEF transcription units are stored. Making the Part-library is the first step. The second step is to outline how the plasmid should look in the end. Here, the mid-copy origin p15A flanked by an ampicillin resistance cassette and a transcription unit, both pointing away from the ori, shall be built. In the next step, the MCF-Positions can be linearly displayed in the arrangement from the outline using a cloning software. Now, the motifs that flank the MCF-Positions have to be linked by MCF-Linkers. There are MCF-Linkers for every combination. Pick the ones that have both motifs to be bridged in their name. The last step is the One-Pot Reaction, as described in the methods.

Bibliography

S. K. Amundsen, A. F. Taylor, A. M. Chaudhury, and G. R. Smith. recD: the gene for an essential third subunit of exonuclease V. Proceedings of the National Academy of Sciences of the United States of America, 83 (15):5558–62, aug 1986. ISSN 0027-8424. doi: 10.1073/pnas.83.15.5558. URL http://www.ncbi.nlm.nih.gov/pubmed/3526335http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC386327.

T. Aoyama and M. Takanami. Supercoiling response of E. coli promoters with different spacer lengths. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression, 949(3):311–317, mar 1988. ISSN 01674781. doi: 10.1016/0167-4781(88)90157-1. URL https://linkinghub.elsevier.com/retrieve/pii/0167478188901571.

A. Balandina, L. Claret, R. Hengge-Aronis, and J. Rouviere-Yaniv. The Escherichia coli histone-like protein HU regulates rpoS translation. Molecular Microbiology, 39(4):1069–1079, dec 2001. ISSN 0950382X. doi: 10.1046/j.1365-2958.2001.02305.x. URL http://doi.wiley.com/10.1046/j.1365-2958.2001.02305.x.

P. Bernard, K. E. Kézdy, L. Van Melderen, J. Steyaert, L. Wyns, M. L. Pato, P. N. Higgins, and M. Couturier. The F Plasmid CcdB Protein Induces Efficient ATP-dependent DNA Cleavage by Gyrase, 1993. ISSN 00222836. URL http://linkinghub.elsevier.com/retrieve/pii/S0022283683716098.

I. Bervoets and D. Charlier. Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. FEMS Microbiology Reviews, 43(3):304–339, may 2019. ISSN 1574-6976. doi: 10.1093/femsre/fuz001. URL https://academic.oup.com/femsre/article/43/3/304/5306444.

H. M. Beyer, P. Gonschorek, S. L. Samodelov, M. Meier, W. Weber, and M. D. Zurbriggen. AQUA Cloning: A Versatile and Simple Enzyme-Free Cloning Approach. PLOS ONE, 10(9):e0137652, sep 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0137652. URL http://dx.plos.org/10.1371/journal.pone.0137652.

J. Blazeck and H. S. Alper. Promoter engineering: Recent advances in controlling transcription at the most fundamental level. Biotechnology Journal, 8(1):46–58, jan 2013. ISSN 18606768. doi: 10.1002/biot.201200120. URL http://doi.wiley.com/10.1002/biot.201200120.

N. Blot, R. Mavathur, M. Geertz, A. Travers, and G. Muskhelishvili. Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. EMBO reports, 7(7):710–5, 2006. ISSN 1469-221X. doi: 10.1038/sj.embor.7400729. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1500834{%}7B{&}{%}7Dtool=pmcentrez{%}7B{&}{%}7Drendertype=abstract{%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1500834{&}tool=pmcentrez{&}rendertype=abstract.

C. E. Bogden, D. Fass, N. Bergman, M. D. Nichols, and J. M. Berger. The structural basis for terminator recognition by the Rho transcription termination factor. Molecular cell, 3(4):487–93, apr 1999. ISSN 1097-2765. doi: 10.1016/s1097-2765(00)80476-1. URL http://www.ncbi.nlm.nih.gov/pubmed/10230401.

B. Bolognesi and B. Lehner. Reaching the limit. eLife, 7, aug 2018. ISSN 2050-084X. doi: 10.7554/eLife.39804. URL https://elifesciences.org/articles/39804.

S. Brenner, F. Jacob, and M. Meselson. An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. Nature, 190(4776):576–581, may 1961. ISSN 0028-0836. doi: 10.1038/190576a0. URL http://www.nature.com/articles/190576a0.

D. F. Browning and S. J. W. Busby. The regulation of bacterial transcription initiation. Nature Reviews Microbiology, 2(1):57–65, jan 2004. ISSN 1740-1526. doi: 10.1038/nrmicro787. URL http://www.nature.com/articles/nrmicro787.

T. Burr, J. Mitchell, A. Kolb, S. Minchin, and S. Busby. DNA sequence elements located immediately upstream of the -10 hexamer in Escherichia coli promoters:

a systematic study. Nucleic Acids Research, 28(9):1864–1870, may 2000. ISSN 0305-1048. doi: 10.1093/nar/28.9.1864. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/28.9.1864.

M. J. Casadaban and S. N. Cohen. Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. Journal of molecular biology, 138(2):179–207, apr 1980. ISSN 0022-2836. URL http://www.ncbi.nlm.nih.gov/pubmed/6997493.

A. Casini, M. Storch, G. S. Baldwin, and T. Ellis. Bricks and blueprints: methods and standards for DNA assembly. Nature Reviews Molecular Cell Biology, 16(9): 568–576, jun 2015. ISSN 1471-0072. doi: 10.1038/nrm4014. URL http://www.nature.com/doifinder/10.1038/nrm4014.

J. J. Champoux. DNA topoisomerases: structure, function, and mechanism. Annual review of biochemistry, 70:369–413, 2001. ISSN 0066-4154. doi: 10.1146/annurev.biochem.70.1.369. URL http://www.ncbi.nlm.nih.gov/pubmed/11395412.

X. Chen, R. Liu, Z. Ma, X. Xu, H. Zhang, J. Xu, Q. Ouyang, and Y. Yang. An extraordinary stringent and sensitive light-switchable gene expression system for bacterial cells. Cell research, 26(7):854–7, 2016. ISSN 1748-7838. doi: 10.1038/cr.2016.74. URL http://www.ncbi.nlm.nih.gov/pubmed/27311594http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5129885.

S. Chong, C. Chen, H. Ge, and X. S. Xie. Mechanism of transcriptional bursting in bacteria. Cell, 158(2):314–326, jul 2014. ISSN 1097-4172. doi: 10.1016/j.cell.2014.05.038. URL http://www.ncbi.nlm.nih.gov/pubmed/25036631http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4105854.

F. Crick. Central dogma of molecular biology. Nature, 227(5258):561–3, aug 1970. ISSN 0028-0836. doi: 10.1038/227561a0. URL http://www.ncbi.nlm.nih.gov/pubmed/4913914.

S. Dages, K. Dages, X. Zhi, and F. Leng. Inhibition of the gyrA promoter by transcription-coupled DNA supercoiling in Escherichia coli. Scientific Reports, 8(1):1–9, 2018. ISSN 20452322. doi: 10.1038/s41598-018-33089-4. URL http://dx.doi.org/10.1038/s41598-018-33089-4.

M. C. Davis, C. A. Kesthely, E. A. Franklin, and S. R. MacLellan. The essential activities of the bacterial sigma factor. Canadian Journal of Microbiology, 63(2): 89–99, feb 2017. ISSN 0008-4166. doi: 10.1139/cjm-2016-0576. URL http://www.nrcresearchpress.com/doi/10.1139/cjm-2016-0576.

C. G. de Boer, E. D. Vaishnav, R. Sadeh, E. L. Abeyta, N. Friedman, and A. Regev. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. Nature Biotechnology, 38(1):56–65, 2020. ISSN 15461696. doi: 10.1038/s41587-019-0315-8.

J. E. Deweese, M. A. Osheroff, and N. Osheroff. DNA Topology and Topoisomerases: Teaching a "Knotty" Subject. Biochemistry and molecular biology education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology, 37(1):2–10, jan 2008. ISSN 1539-3429. doi: 10.1002/bmb.20244. URL http://doi.wiley.com/10.1002/bmb.20244http://www.ncbi.nlm.nih.gov/pubmed/19225573http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2643378.

A. J. Dombroski, W. A. Walter, M. T. Record, D. A. Slegele, and C. A. Gross. Polypeptides containing highly conserved regions of transcription initiation factor σ70 exhibit specificity of binding to promoter DNA. Cell, 70(3):501–512, 1992. ISSN 00928674. doi: 10.1016/0092-8674(92)90174-B.

C. Dorman. Flexible response: DNA supercoiling, transcription and bacterial adaptation to environmental stress. Trends in Microbiology, 4(6):214–216, jun 1996. ISSN 0966842X. doi: 10.1016/0966-842X(96)30015-2. URL https://linkinghub.elsevier.com/retrieve/pii/0966842X96300152.

T. Ellis, T. Adie, and G. S. Baldwin. DNA assembly for synthetic biology: from parts to pathways and beyond. Integrative Biology, 3(2):109, 2011. ISSN 1757-9694. doi: 10.1039/c0ib00070a. URL http://xlink.rsc.org/?DOI=c0ib00070a.

L. Emde. Etablierung eines flexiblen Klonierungssystems zur Untersuchung positionsabhängiger Promotoraktivität, 2017.

C. Engler, R. Kandzia, and S. Marillonnet. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. PLoS ONE, 3(11):e3647, nov 2008. ISSN

1932-6203. doi: 10.1371/journal.pone.0003647. URL http://dx.plos.org/10.1371/journal.pone.0003647.

C. Engler, M. Youles, R. Gruetzner, T.-M. Ehnert, S. Werner, J. D. G. Jones, N. J. Patron, and S. Marillonnet. A Golden Gate Modular Cloning Toolbox for Plants. ACS Synthetic Biology, 3(11):839–843, nov 2014. ISSN 2161-5063. doi: 10.1021/sb4001504. URL http://pubs.acs.org/doi/abs/10.1021/sb4001504.

S. T. Estrem, T. Gaal, W. Ross, and R. L. Gourse. Identification of an UP element consensus sequence for bacterial promoters. Proceedings of the National Academy of Sciences of the United States of America, 95(17):9761–9766, 1998. ISSN 00278424. doi: 10.1073/pnas.95.17.9761.

P. J. Farnham and T. Platt. Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. Nucleic acids research, 9(3):563–77, feb 1981. ISSN 0305-1048. doi: 10.1093/nar/9.3.563. URL http://www.ncbi.nlm.nih.gov/pubmed/7012794http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC327222.

R. D. Finn, E. V. Orlova, B. Gowen, M. Buck, and M. van Heel. Escherichia coli RNA polymerase core and holoenzyme structures. The EMBO journal, 19(24):6833–44, dec 2000. ISSN 0261-4189. doi: 10.1093/emboj/19.24.6833. URL http://www.ncbi.nlm.nih.gov/pubmed/11118218http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC305883.

G. Fulcrand, S. Dages, X. Zhi, P. Chapagain, B. S. Gerstman, D. Dunlap, and F. Leng. DNA supercoiling, a critical signal regulating the basal expression of the lac operon in Escherichia coli. Scientific Reports, 6(October 2015):1–12, 2016. ISSN 20452322. doi: 10.1038/srep19243.

A. Gabrielian, A. Simoncsits, and S. Pongor. Distribution of bending propensity in DNA sequences. FEBS Letters, 393(1):124–130, 1996. ISSN 00145793. doi: 10.1016/0014-5793(96)00837-X.

A. Gabrielian, K. Vlahovicek, and S. Pongor. Distribution of sequence-dependent curvature in genomic DNA sequences. FEBS Letters, 406(1-2):69–74, apr 1997. ISSN

00145793. doi: 10.1016/S0014-5793(97)00236-6. URL http://doi.wiley.com/10.1016/S0014-5793{%}2897{%}2900236-6.

J. Goedhart, D. von Stetten, M. Noirclerc-Savoye, M. Lelimousin, L. Joosen, M. A. Hink, L. van Weeren, T. W. Gadella, and A. Royant. Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. Nature Communications, 3 (1):751, jan 2012. ISSN 2041-1723. doi: 10.1038/ncomms1738. URL http://www.nature.com/articles/ncomms1738.

J. Gowrishankar and R. Harinarayanan. Why is transcription coupled to translation in bacteria? Molecular Microbiology, 54(3):598–603, sep 2004. ISSN 0950382X. doi: 10.1111/j.1365-2958.2004.04289.x. URL http://doi.wiley.com/10.1111/j.1365-2958.2004.04289.x.

R. Green and E. J. Rogers. Transformation of chemically competent E. coli. Methods in enzymology, 529:329–36, 2013. ISSN 1557-7988. doi: 10.1016/B978-0-12-418687-3.00028-8. URL http://www.ncbi.nlm.nih.gov/pubmed/24011059http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4037286.

D. K. Hawley and W. R. McClure. Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic acids research, 11 (8):2237–55, apr 1983. ISSN 0305-1048. doi: 10.1093/nar/11.8.2237. URL http://www.ncbi.nlm.nih.gov/pubmed/6344016http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC325881.

K. L. Henderson, L. C. Felth, C. M. Molzahn, I. Shkel, S. Wang, M. Chhabra, E. F. Ruff, L. Bieter, J. E. Kraft, and M. T. Record. Mechanism of transcription initiation and promoter escape by E. coli RNA polymerase. Proceedings of the National Academy of Sciences of the United States of America, 114(15):E3032–E3040, 2017. ISSN 1091-6490. doi: 10.1073/pnas.1618675114. URL http://www.ncbi.nlm.nih.gov/pubmed/28348246http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5393250.

R. Hengge-Aronis. Signal Transduction and Regulatory Mechanisms Involved in Control of the S (RpoS) Subunit of RNA Polymerase. Microbiology and Molecular Biology Reviews, 66(3):373–395, 2002. ISSN 1092-2172. doi: 10.1128/mmbr.66.3.373-395.2002.

N. P. Higgins. Species-specific supercoil dynamics of the bacterial nucleoid. Biophysical Reviews, 8:113–121, 2016. ISSN 18672469. doi: 10.1007/s12551-016-0207-9. URL http://dx.doi.org/10.1007/s12551-016-0207-9.

H. Hoeksema, J. L. Johnson, and J. W. Hinman. STRUCTURAL STUDIES ON STREPTONIVICIN, A NEW ANTIBIOTIC. Journal of the American Chemical Society, 77(24):6710–6711, dec 1955. ISSN 0002-7863. doi: 10.1021/ja01629a129. URL https://pubs.acs.org/doi/abs/10.1021/ja01629a129.

R. W. Holley. Structure of an alanine transfer ribonucleic acid. JAMA, 194(8):868–71, nov 1965. ISSN 0098-7484. URL http://www.ncbi.nlm.nih.gov/pubmed/5898068.

I. G. Hook-Barnard and D. M. Hinton. Transcription Initiation by Mix and Match Elements: Flexibility for Polymerase Binding to Bacterial Promoters. Gene Regulation and Systems Biology, 1(301):117762500700100, 2007. ISSN 1177-6250. doi: 10.1177/117762500700100020.

M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science, 337(6096):816–821, aug 2012. ISSN 0036-8075. doi: 10.1126/science.1225829. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1225829.

J. Keasling. The Promise of Synthetic Biology. In National Academy of Engineering, editor, Reports on Leading-Edge Engineering from the 2005 Symposium. National Academies Press, Washington, D.C., frontiers edition, feb 2006. ISBN 978-0-309-10102-8. doi: 10.17226/11577. URL http://www.nap.edu/catalog/11577.

C. A. Klein, L. Emde, A. Kuijpers, and P. Sobetzko. MoCloFlex: A Modular Yet Flexible Cloning System. Frontiers in Bioengineering and Biotechnology, 7 (October):1–9, oct 2019. ISSN 2296-4185. doi: 10.3389/fbioe.2019.00271. URL https://www.frontiersin.org/article/10.3389/fbioe.2019.00271/full.

C. A. Klein, M. Teufel, and P. Sobetzko. CRISPR Swap n' Drop: a Modular Genome Editing Tool. in preparation, 2020a.

C. A. Klein, C. J. Weile, and P. Sobetzko. The Spacer of Escherichia coli Promoters Modulates Transcription by Length, Sequence, and sequential TG-Motifs. in preparation, 2020b.

F. Kouzine, J. Liu, S. Sanford, H.-J. Chung, and D. Levens. The dynamic response of upstream DNA to transcription-generated torsional stress. Nature Structural & Molecular Biology, 11(11):1092–1100, nov 2004. ISSN 1545-9993. doi: 10.1038/nsmb848. URL http://www.nature.com/articles/nsmb848.

F. Kouzine, S. Sanford, Z. Elisha-Feil, and D. Levens. The functional response of upstream DNA to dynamic supercoiling in vivo. Nature Structural & Molecular Biology, 15(2):146–154, feb 2008. ISSN 1545-9993. doi: 10.1038/nsmb.1372. URL http://www.nature.com/articles/nsmb.1372.

F. Kouzine, A. Gupta, L. Baranello, D. Wojtowicz, K. Benaissa, J. Liu, T. M. Przytycka, D. Levens, and N. Struct Mol Biol Author manuscript. Transcription dependent dynamic supercoiling is a short-range genomic force HHS Public Access Author manuscript. Nat Struct Mol Biol, 20(3):396–403, 2013. doi: 10.1038/nsmb.2517. URL http://www.nature.com/authors/editorial{_}policies/license.html{#}terms.

J. A. Kraemer, A. G. Sanderlin, and M. T. Laub. The stringent response inhibits DNA replication initiation in E. Coli by modulating supercoiling of oric. mBio, 10(4):1–18, 2019. ISSN 21507511. doi: 10.1128/mBio.01330-19.

G.-J. Kremers, J. Goedhart, E. B. van Munster, and T. W. J. Gadella. Cyan and Yellow Super Fluorescent Proteins with Improved Brightness, Protein Folding, and FRET Förster Radius † , ‡. Biochemistry, 45(21):6570–6580, may 2006. ISSN 0006-2960. doi: 10.1021/bi0516273. URL https://pubs.acs.org/doi/10.1021/bi0516273.

A. Kumar, R. A. Malloch, N. Fujita, D. A. Smillie, A. Ishihama, and R. S. Hayward. The Minus 35-Recognition Region of Escherichia coli Sigma 70 is Inessential for Initiation of Transcription at an "Extended Minus 10" Promoter. Journal of Molecular Biology, 232(2):406–418, jul 1993. ISSN 00222836. doi: 10.1006/jmbi.1993.1400. URL https://linkinghub.elsevier.com/retrieve/pii/S0022283683714002.

A. Lal, A. Dhar, A. Trostel, F. Kouzine, A. S. N. Seshasayee, and S. Adhya. Genome scale patterns of supercoiling in a bacterial chromosome. Nature communications, 7:11055, mar 2016. ISSN 2041-1723. doi: 10.1038/ncomms11055. URL http://www.ncbi.nlm.nih.gov/pubmed/27025941http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4820846.

M. E. Lee, W. C. DeLoache, B. Cervantes, and J. E. Dueber. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synthetic Biology, 4(9):975–986, sep 2015. ISSN 2161-5063. doi: 10.1021/sb500366v. URL http://pubs.acs.org/doi/10.1021/sb500366v.

D. M. J. Lilley and C. F. Higgins. Local DNA topology and gene expression: the case of the leu-500 promoter. Molecular Microbiology, 5(4):779–783, apr 1991. ISSN 0950382X. doi: 10.1111/j.1365-2958.1991.tb00749.x. URL http://doi.wiley.com/10.1111/j.1365-2958.1991.tb00749.x.

H. M. Lim, D. E. A. Lewis, H. J. Lee, M. Liu, and S. Adhya. Effect of Varying the Supercoiling of DNA on Transcription and Its Regulation. Biochemistry, 42(36): 10718–10725, sep 2003. ISSN 0006-2960. doi: 10.1021/bi030110t. URL https://pubs.acs.org/doi/10.1021/bi030110t.

L. F. Liu and J. C. Wang. Supercoiling of the DNA template during transcription. Proceedings of the National Academy of Sciences, 84(20):7024–7027, 1987. ISSN 0027-8424. doi: 10.1073/pnas.84.20.7024. URL http://www.pnas.org/cgi/doi/10.1073/pnas.84.20.7024.

J. Ma and M. D. Wang. DNA supercoiling during transcription. Biophysical Reviews, 8:75–87, 2016. ISSN 18672469. doi: 10.1007/s12551-016-0215-9. URL http://dx.doi.org/10.1007/s12551-016-0215-9.

A. Maxwell. The interaction between coumarin drugs and DNA gyrase. Molecular microbiology, 9(4):681–6, aug 1993. ISSN 0950-382X. doi: 10.1111/j.1365-2958.1993.tb01728.x. URL http://www.ncbi.nlm.nih.gov/pubmed/8231802.

A. McCracken and P. Timms. Efficiency of transcription from promoter sequence variants in Lactobacillus is both strain and context dependent. Journal of Bacteriology, 181(20):6569–6572, 1999. ISSN 00219193. doi: 10.1128/jb.181.20.6569-6572.1999.

S. J. Messerschmidt, D. Schindler, C. M. Zumkeller, F. S. Kemter, N. Schallopp, and T. Waldminghaus. Optimization and Characterization of the Synthetic Secondary Chromosome synVicII in Escherichia coli. Frontiers in Bioengineering and Biotechnology, 4, dec 2016. ISSN 2296-4185. doi: 10.3389/fbioe.2016.00096. URL http://journal.frontiersin.org/article/10.3389/fbioe.2016.00096/full.

S. Meyer and G. Beslon. Torsion-Mediated Interaction between Adjacent Genes. PLoS Computational Biology, 10(9), 2014. ISSN 15537358. doi: 10.1371/journal.pcbi.1003785.

P. Mittal, J. Brindle, J. Stephen, J. B. Plotkin, and G. Kudla. Codon usage influences fitness through RNA toxicity. Proceedings of the National Academy of Sciences, 115(34):8639–8644, aug 2018. ISSN 0027-8424. doi: 10.1073/pnas.1810022115. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1810022115.

C. Monteilhet, A. Perrin, A. Thierry, L. Colleaux, and B. Dujon. Purification and characterization of the in vitro activity of I- Sce I, a novel and highly specific endonuclease encoded by a group I intron. Nucleic Acids Research, 18(6):1407–1413, 1990. ISSN 0305-1048. doi: 10.1093/nar/18.6.1407. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/18.6.1407.

S. J. Moore, H. E. Lai, R. J. R. Kelwick, S. M. Chee, D. J. Bell, K. M. Polizzi, and P. S. Freemont. EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology. ACS Synthetic Biology, 5(10):1059–1069, 2016. ISSN 21615063. doi: 10.1021/acssynbio.6b00031.

K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich. Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. Cold Spring Harbor Symposia on Quantitative Biology, 51:263–273, jan 1986. ISSN 0091-7451. doi: 10.1101/SQB.1986.051.01.032. URL http://symposium.cshlp.org/cgi/doi/10.1101/SQB.1986.051.01.032.

G. Muskhelishvili and A. Travers. Transcription factor as a topological homeostat. Frontiers in bioscience : a journal and virtual library, 8:d279–85, jan 2003. ISSN 1093-9946. doi: 10.2741/969. URL http://www.ncbi.nlm.nih.gov/pubmed/12456360.

G. Muskhelishvili, A. Travers, and A. Travers. The regulatory role of DNA supercoiling in nucleoprotein complex assembly and genetic activity. Biophysical Reviews, 8, 2016. ISSN 1867-2450. doi: 10.1007/s12551-016-0237-3. URL http://dx.doi.org/10.1007/s12551-016-0237-3.

S. Nozaki and H. Niki. Exonuclease III (XthA) Enforces In Vivo DNA Cloning of Escherichia coli To Create Cohesive Ends. Journal of Bacteriology, 201(5):e00660–18, mar 2019. doi: 10.1128/JB.00660-18. URL http://jb.asm.org/content/201/5/e00660-18.abstract.

T. Ó Cróinín, R. K. Carroll, A. Kelly, and C. J. Dorman. Roles for DNA supercoiling and the Fis protein in modulating expression of virulence genes during intracellular growth of Salmonella enterica serovar Typhimurium. Molecular Microbiology, 62 (3):869–882, 2006. ISSN 0950382X. doi: 10.1111/j.1365-2958.2006.05416.x.

P. Oldham, S. Hall, and G. Burton. Synthetic Biology: Mapping the Scientific Landscape. PLoS ONE, 7(4):e34368, apr 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0034368. URL https://dx.plos.org/10.1371/journal.pone.0034368.

L. Ortiz, M. Pavan, L. McCarthy, J. Timmons, and D. M. Densmore. Automated Robotic Liquid Handling Assembly of Modular DNA Devices. Journal of Visualized Experiments, (130), dec 2017. ISSN 1940-087X. doi: 10.3791/54703. URL https://www.jove.com/video/54703/automated-robotic-liquid-handling-assembly-of-modular-dna-devices.

Y. H. Park, C. R. Lee, M. Choe, and Y. J. Seok. HPr antagonizes the anti-σ70 activity of Rsd in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America, 110(52):21142–21147, 2013. ISSN 00278424. doi: 10.1073/pnas.1316629111.

A. S. Rajkumar, J. A. Varela, H. Juergens, J.-M. G. Daran, and J. P. Morrissey. Biological Parts for Kluyveromyces marxianus Synthetic Biology. Frontiers in Bioengineering and Biotechnology, 7, may 2019. ISSN 2296-4185. doi: 10.3389/fbioe.2019.00097. URL https://www.frontiersin.org/article/10.3389/fbioe.2019.00097/full.

C. R. Reisch and K. L. J. Prather. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli. Scientific reports, 5:15096, oct 2015. ISSN 2045-2322. doi: 10.1038/srep15096. URL http://www.ncbi.nlm.nih.gov/pubmed/26463009http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4604488.

V. Rifka, V. D. Jaco, and D. Cees. Counterintuitive DNA sequence dependence in supercoiling-induced DNA melting. PLoS ONE, 10(10):1–14, 2015. ISSN 19326203. doi: 10.1371/journal.pone.0141576.

C. W. Roberts and J. W. Roberts. Base-specific recognition of the nontemplate strand of promoter DNA by E. coli RNA polymerase. Cell, 86(3):495–501, 1996. ISSN 00928674. doi: 10.1016/S0092-8674(00)80122-1.

W. Ross, K. K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov, and R. L. Gourse. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science (New York, N.Y.), 262(5138): 1407–13, nov 1993. ISSN 0036-8075. doi: 10.1126/science.8248780. URL http://www.ncbi.nlm.nih.gov/pubmed/8248780.

RStudio Team. RStudio: Integrated Development Environment for R. RStudio, Inc., Boston, MA, 2015. URL http://www.rstudio.com/.

R. M. Saecker, M. T. Record, and P. L. DeHaseth. Mechanism of Bacterial Transcription Initiation: RNA Polymerase - Promoter Binding, Isomerization to Initiation-Competent Open Complexes, and Initiation of RNA Synthesis. Journal of Molecular Biology, 412(5):754–771, oct 2011. ISSN 00222836. doi: 10.1016/j.jmb.2011.01.018. URL https://linkinghub.elsevier.com/retrieve/pii/S0022283611000350.

H. Salgado, I. Martínez-Flores, V. H. Bustamante, K. Alquicira-Hernández, J. S. García-Sotelo, D. García-Alonso, and J. Collado-Vides. Using RegulonDB, the Escherichia coli K-12 Gene Regulatory Transcriptional Network Database. Current protocols in bioinformatics, 61(1):1.32.1–1.32.30, 2018. ISSN 1934-340X. doi: 10.1002/cpbi.43. URL http://www.ncbi.nlm.nih.gov/pubmed/30040192http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6060643.

J. SantaLucia. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proceedings of the National Academy of Sciences of the United States of America, 95(4):1460–5, feb 1998. ISSN 0027-8424. doi: 10.1073/pnas.95.4.1460. URL http://www.ncbi.nlm.nih.gov/pubmed/9465037http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC19045.

A. Sarrion-Perdigones, M. Vazquez-Vilar, J. Palaci, B. Castelijns, J. Forment, P. Ziarsolo, J. Blanca, A. Granell, and D. Orzaez. GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. PLANT PHYSIOLOGY, 162 (3):1618–1631, jul 2013. ISSN 0032-0889. doi: 10.1104/pp.113.217661. URL http://www.plantphysiol.org/cgi/doi/10.1104/pp.113.217661.

D. Schindler, S. Milbredt, T. Sperlea, and T. Waldminghaus. Design and Assembly of DNA Sequence Libraries for Chromosomal Insertion in Bacteria Based on a Set of Modified MoClo Vectors. ACS Synthetic Biology, 5(12):1362–1368, dec 2016. ISSN 2161-5063. doi: 10.1021/acssynbio.6b00089. URL http://pubs.acs.org/doi/abs/10.1021/acssynbio.6b00089.

R. Schneider, A. Travers, T. Kutateladze, and G. Muskhelishvili. A DNA architectural protein couples cellular physiology and DNA topology in Escherichia coli. Molecular microbiology, 34(5):953–64, dec 1999. ISSN 0950-382X. doi: 10.1046/j.1365-2958.1999.01656.x. URL http://www.ncbi.nlm.nih.gov/pubmed/10594821.

A. Schrauwers and B. Poolman. Synthetische Biologie. Springer Berlin Heidelberg, Berlin, Heidelberg, jun 2013. ISBN 978-3-642-34592-0. doi: 10.1007/978-3-642-34593-7. URL http://link.springer.com/10.1007/s12268-011-0063-xhttp://link.springer.com/10.1007/978-3-642-34593-7.

W. Schröder, J. Bernhardt, G. Marincola, L. Klein-Hitpass, A. Herbig, G. Krupp, K. Nieselt, and C. Wolz. Altering gene expression by aminocoumarins: the role of DNA supercoiling in Staphylococcus aureus. BMC genomics, 15: 291, apr 2014. ISSN 1471-2164. doi: 10.1186/1471-2164-15-291. URL http://www.ncbi.nlm.nih.gov/pubmed/24734910http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4023603.

G. Selzer, T. Som, T. Itoh, and J.-i. Tomizawa. The origin of replication of plasmid p15A and comparative studies on the nucleotide sequences around the origin of related plasmids. Cell, 32(1):119–129, jan 1983. ISSN 00928674. doi: 10.1016/0092-8674(83)90502-0. URL https://linkinghub.elsevier.com/retrieve/pii/0092867483905020.

X. Shu, N. C. Shaner, C. A. Yarbrough, R. Y. Tsien, and S. J. Remington. Novel Chromophores and Buried Charges Control Color in mFruits † , ‡. Biochemistry,

45(32):9639–9647, aug 2006. ISSN 0006-2960. doi: 10.1021/bi060773l. URL https://pubs.acs.org/doi/10.1021/bi060773l.

R. K. Shultzaberger, Z. Chen, K. A. Lewis, and T. D. Schneider. Anatomy of Escherichia coli σ 70 promoters. Nucleic Acids Research, 35(3):771–788, feb 2007. ISSN 1362-4962. doi: 10.1093/nar/gkl956. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkl956.

P. Sobetzko. Transcription-coupled DNA supercoiling dictates the chromosomal arrangement of bacterial genes. Nucleic Acids Research, 44(4):1514–1524, 2016. ISSN 13624962. doi: 10.1093/nar/gkw007.

P. Sobetzko, M. Glinkowska, A. Travers, and G. Muskhelishvili. DNA thermodynamic stability and supercoil dynamics determine the gene expression program during the bacterial growth cycle. Molecular bioSystems, 9(7):1643–51, jul 2013. ISSN 1742-2051. doi: 10.1039/c3mb25515h. URL http://www.ncbi.nlm.nih.gov/pubmed/23493878.

B. Spira, X. Hu, and T. Ferenci. Strain variation in ppGpp concentration and RpoS levels in laboratory strains of Escherichia coli K-12. Microbiology, 154(9):2887–2895, 2008. ISSN 13500872. doi: 10.1099/mic.0.2008/018457-0.

D. M. Stoebel, A. M. Dean, and D. E. Dykhuizen. The Cost of Expression of Escherichia coli lac Operon Proteins Is in the Process, Not in the Products. Genetics, 178(3):1653–1660, mar 2008. ISSN 0016-6731. doi: 10.1534/genetics.107.085399. URL http://www.genetics.org/lookup/doi/10.1534/genetics.107.085399.

M. Storch, A. Casini, B. Mackrow, T. Fleming, H. Trewhitt, T. Ellis, and G. S. Baldwin. BASIC: A New Biopart Assembly Standard for Idempotent Cloning Provides Accurate, Single-Tier DNA Assembly for Synthetic Biology. ACS Synthetic Biology, 4(7):781–787, jul 2015. ISSN 21615063. doi: 10.1021/sb500356d. URL http://pubs.acs.org/doi/pdf/10.1021/sb500356dhttp://pubs.acs.org/doi/abs/10.1021/sb500356d.

D. Stueber and H. Bujard. Transcription from efficient promoters can interfere with plasmid replication and diminish expression of plasmid specified genes. The EMBO Journal, 1(11):1399–1404, nov 1982. ISSN 02614189. doi: 10.1002/j.1460-2075.1982.tb01329.x. URL http://doi.wiley.com/10.1002/j.1460-2075.1982.tb01329.x.

L. C. Thomason, J. A. Sawitzke, X. Li, N. Costantino, and D. L. Court. Recombineering: genetic engineering in bacteria using homologous recombination. Current protocols in molecular biology, 106:1.16.1–39, apr 2014. ISSN 1934-3647. doi: 10.1002/0471142727.mb0116s106. URL http://www.ncbi.nlm.nih.gov/pubmed/24733238.

A. Travers and G. Muskhelishvili. DNA supercoiling — a global transcriptional regulator for enterobacterial growth? Nature Reviews Microbiology, 3(2):157–169, feb 2005. ISSN 1740-1526. doi: 10.1038/nrmicro1088. URL http://www.nature.com/articles/nrmicro1088.

H. Urlaub, V. Kruft, O. Bischof, E. C. Müller, and B. Wittmann-Liebold. Protein-rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies. The EMBO journal, 14(18):4578–88, sep 1995. ISSN 0261-4189. URL http://www.ncbi.nlm.nih.gov/pubmed/7556101http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC394550.

B. Van Hove, C. Guidi, L. De Wannemaeker, J. Maertens, and M. De Mey. Recursive DNA Assembly Using Protected Oligonucleotide Duplex Assisted Cloning (PODAC). ACS Synthetic Biology, 6(6):943–949, jun 2017. ISSN 2161-5063. doi: 10.1021/acssynbio.7b00017. URL http://pubs.acs.org/doi/abs/10.1021/acssynbio.7b00017.

S. Vecchione and G. Fritz. CRIMoClo plasmids for modular assembly and orthogonal chromosomal integration of synthetic circuits in Escherichia coli. Journal of Biological Engineering, 13(1):92, dec 2019. ISSN 1754-1611. doi: 10.1186/s13036-019-0218-8. URL https://jbioleng.biomedcentral.com/articles/10.1186/s13036-019-0218-8.

J. C. Wang. Helical repeat of DNA in solution. Proceedings of the National Academy of Sciences, 76(1):200–203, jan 1979. ISSN 0027-8424. doi: 10.1073/pnas.76.1.200. URL http://www.pnas.org/cgi/doi/10.1073/pnas.76.1.200.

J. D. Watson and F. H. Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature, 171(4356):737–8, apr 1953. ISSN 0028-0836. doi: 10.1038/171737a0. URL http://www.ncbi.nlm.nih.gov/pubmed/13054692.

E. Weber, C. Engler, R. Gruetzner, S. Werner, and S. Marillonnet. A Modular Cloning System for Standardized Assembly of Multigene Constructs. PLoS ONE, 6(2): e16765, feb 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0016765. URL http://dx.plos.org/10.1371/journal.pone.0016765.

C. J. Weile. Ein optogenetischer Versuchsaufbau zur Untersuchung synthetischer supercoiling-sensitiver Sigma-70-Promotoren, 2018.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL https://ggplot2.tidyverse.org.

K. S. Wilson and P. H. von Hippel. Transcription termination at intrinsic terminators: the role of the RNA hairpin. Proceedings of the National Academy of Sciences of the United States of America, 92(19):8793–7, sep 1995. ISSN 0027-8424. doi: 10.1073/pnas.92.19.8793. URL http://www.ncbi.nlm.nih.gov/pubmed/7568019http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC41053.

Y. Yamaguchi. Transcription Elongation, Mechanisms, pages 2221–2224. Springer New York, New York, NY, 2013. ISBN 978-1-4419-9863-7. doi: 10.1007/978-1-4419-9863-7_1407. URL https://doi.org/10.1007/978-1-4419-9863-7{_}1407.

M. Yu, J. Souaya, and D. A. Julin. The 30-kDa C-terminal domain of the RecB protein is critical for the nuclease activity, but not the helicase activity, of the RecBCD enzyme fromEscherichiacoli. Proceedings of the National Academy of Sciences, 95 (3):981–986, feb 1998. ISSN 0027-8424. doi: 10.1073/PNAS.95.3.981. URL https://www.pnas.org/content/95/3/981.

X. Zhi, S. Dages, K. Dages, Y. Liu, Z. C. Hua, J. Makemson, and F. Leng. Transient and dynamic DNA supercoiling potently stimulates the leu-500 promoter in Escherichia coli. Journal of Biological Chemistry, 292(35):14566–14575, 2017. ISSN 1083351X. doi: 10.1074/jbc.M117.794628.

C. Zumkeller, D. Schindler, and T. Waldminghaus. Modular Assembly of Synthetic Secondary Chromosomes. In R. T. Dame, editor, Bacterial Chromatin: Methods

and Protocols, pages 71–94. Springer New York, New York, NY, 2018. ISBN 978-1-4939-8675-0. doi: 10.1007/978-1-4939-8675-0_5. URL https://doi.org/10.1007/978-1-4939-8675-0{_}5.

Curriculum Vitae

Danksagung

Diese Arbeit war mir auch deshalb möglich, weil ich das Glück habe die richtigen Menschen um mich zu wissen. Deshalb möchte ich an dieser Stelle danke sagen.

Meiner Familie, insbesondere meinen Eltern, möchte ich danken. Ohne Euch wäre ich nicht, wo ich heute bin. Meiner Frau Janina kann ich an dieser Stelle gar nicht genug danken – Worte können nur schwer beschreiben was mir unsere Ehe bedeutet. Meinen Freunden möchte ich an dieser Stelle dafür danke sagen, dass sie immer für mich da sind.

Bei meinem Doktorvater Dr. Patrick Sobetzko möchte ich mich dafür bedanken, dass er mir die Möglichkeit zur Promotion gegeben hat. Außerdem für all die guten Gespräche, seine Geduld bei meinen ersten Programmierschritten (1 "das i"), den zahlosen Ratschlägen und den ausgiebigen Diskussionen zu allen (!) möglichen Themen. Prof. Dr. Torsten Waldminghaus gilt mein Dank dafür, dass er mich in seinem Labor aufgenommen hat und dass er bei Bedarf immer die richtigen Worte gefunden hat. Bei Prof. Dr. Stefan Rensing und Prof. Dr. Martin Thanbichler möchte ich mich für ihre Prüfertätigkeit bedanken.

Meinen Kollegen der letzten vier Jahren gilt mein Dank dafür, dass ich in einer wirklich geselligen und freundschaftlichen Umgebung arbeiten durfte. Danke an Dr. Franziska Kemter, Celine Zumkeller, Marc Teufel und David Velázquez sowie allen Studenten der AG Waldminghaus. Ein besonderes Dankeschön auch an meine ehemaligen Studenten Aaron Kuijpers, Carl Weile und Leonie Emde, die Zusammenarbeit mit euch war klasse. Bei Beatrice Schofield möchte ich mich an dieser Stelle für ihre hilfreichen Hinweise zur englischen Sprache bedanken.

Dr. Georg Fritz und seiner AG möchte ich danke sagen, dafür dass ich mich jeden Tag auf 11:30 Uhr und unser gemeinsames Mittagessen gefreut habe.

Den Mitarbeitern der AG Becker, der Geschäftsstelle von Synmikro und nicht zuletzt Mona Bastian gilt mein Dank für die reibungslose Zusammenarbeit auf einer Etage.

An dieser Stelle möchte ich mich auch bei Prof. Dr. Wolfgang Nellen, Dr. Christoph Schiklenk, Dr. Doreen Meier und Dr. Michael Friedrich bedanken. Danke dafür, dass ihr mir eure Begeisterung an der Biologie weitergegeben habt und schon seit Jahren immer ein Ohr sowie den ein oder anderen Ratschlag für mich hattet.

Erklärung der selbstständigen Erarbeitung der Dissertation

Hiermit erkläre ich, dass ich die vorliegende Dissertation:

„An Experimental Framework to Examine the Influence of Promoter Architecture and Genomic Context on Gene Expression“

selbstständig und ohne unerlaubte Hilfsmittel angefertigt habe. Es wurden keine anderen, als der von mir ausdrücklich angegebenen Hilfsmittel verwendet. Die Dissertation wurde in der jetzigen oder in einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht.

Es handelt sich bei den heute von mir eingereichten Exemplaren um in Wort und Bild völlig übereinstimmende Exemplare.

Des Weiteren erkläre ich, dass die digitalen Abbildungen ausschließlich originale Daten enthalten und in keinem Fall inhaltsverändernde Bildbearbeitung vorgenommen wurde.

Marburg, den

_____

Carlo Anton Klein

Publications From This Work

Published Manuscripts

C. A. Klein, L. Emde, A. Kuijpers, and P. Sobetzko. MoCloFlex: A Modular Yet Flexible Cloning System. Frontiers in Bioengineering and Biotechnology, 7(October):1–9, oct 2019. ISSN 2296-4185. doi: 10.3389/fbioe.2019.00271. URL `https://www.frontiersin.org/article/10.3389/fbioe.2019.00271/full`

Parts of this publication are used as the subsections 1.4.1 and 2.1.

Manuscripts In Preparation

C. A. Klein, C. J. Weile, and P. Sobetzko. The Spacer of Escherichia coli Promoters Modulates Transcription by Length, Sequence, and sequential TG-Motifs. in preparation, 2020b

C. A. Klein, M. Teufel, and P. Sobetzko. CRISPR Swap n' Drop: a Modular Genome Editing Tool. in preparation, 2020a